Computational social science

Social networks: Structure

Midjourney:
https://creativecommons.org/licenses/by-nc/4.0/legalcode

Course structure

Period IV

Week	Lecture	Exer. dl	Ext. dl	Topic
1	~~Feb 27~~	~~Mar 3~~	~~Mar 15~~	~~Introduction to CSS~~
2	~~Mar 6~~	~~Mar 10~~	~~Mar 22~~	~~Artificial societies & agent-based models~~
3	~~Mar 13~~	~~Mar 17~~	Mar 29	~~Data & digital traces~~
4	~~Mar 20~~	~~Mar 24~~	Apr 5	~~Counting things & analysing text~~
5	Mar 27	Mar 31	Apr 12	Social networks: structure
6	Apr 3	*	-	Introduction to the project

***Project deadline: May 26**
Project peer review: June 2

Period V

Week	Lecture	Exercise dl	Ext. dl	Topic
7	Apr 24	May 5	May 10	Ethics, privacy, legal
-	-	-	-	WAPPU
8	May 8	May 12**	May 24	Agent-based models & emergence
9	May 15	May 19***	May 31	Social networks: dynamics
10	May 22	May 26***	June 7	Experiments & interventions at scale
11	May 29	-	-	Computing for social good

****Bonus round**
*****Only lecture questions**

SOCIAL SCIENCE

Computational Social Science

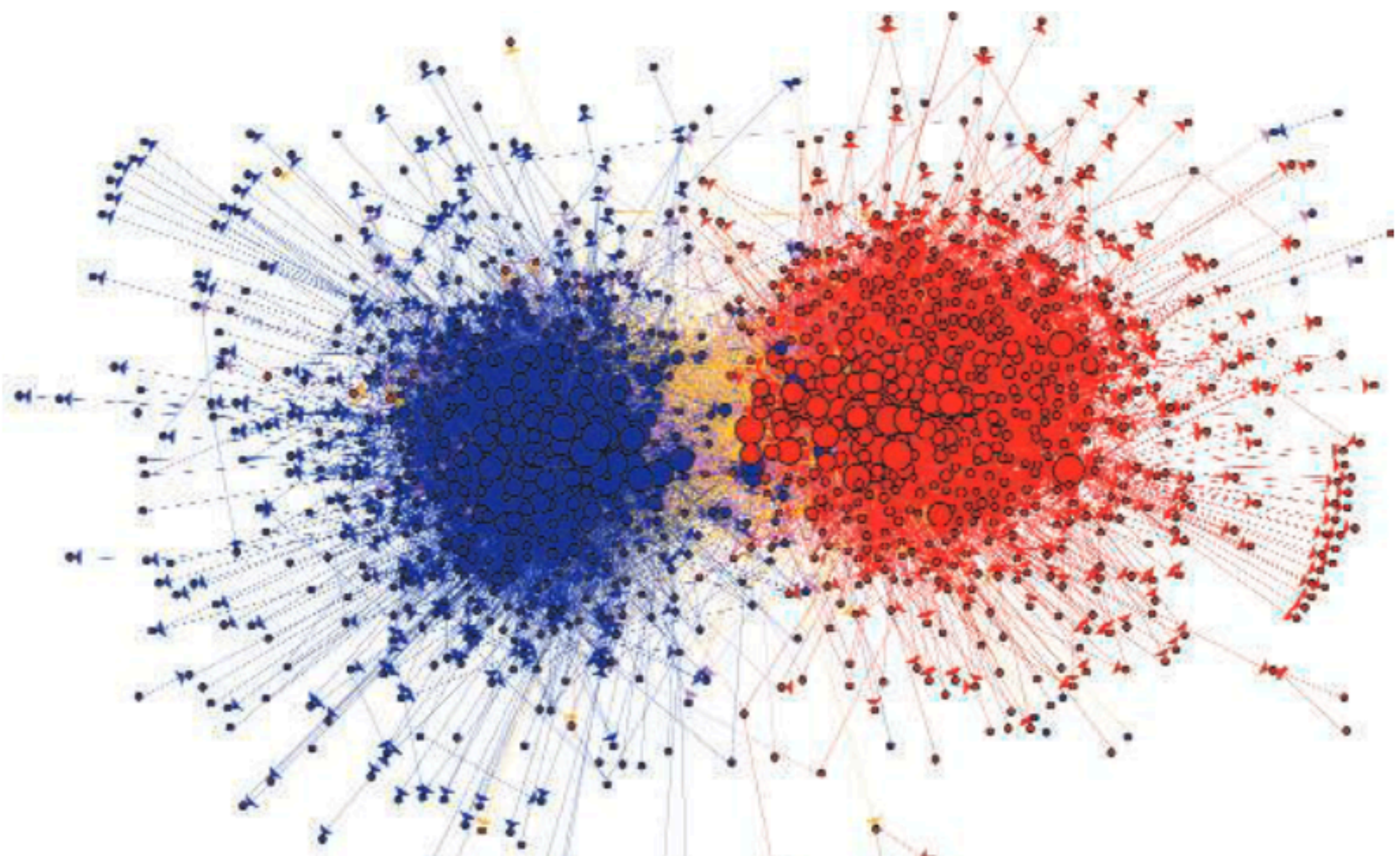
A field is emerging that leverages the capacity to collect and analyze data at a scale that may reveal patterns of individual and group behaviors.

David Lazer,[1] Alex Pentland,[2] Lada Adamic,[3] Sinan Aral,[2,4] Albert-László Barabási,[5] Devon Brewer,[6] Nicholas Christakis,[1] Noshir Contractor,[7] James Fowler,[8] Myron Gutmann,[3] Tony Jebara,[9] Gary King,[1] Michael Macy,[10] Deb Roy,[2] Marshall Van Alstyne[2,11]

We live life in the network. We check our e-mails regularly, make mobile phone calls from almost any location, swipe transit cards to use public transportation, and make purchases with credit cards. Our movements in public places may be captured by video cameras, and our medical records stored as digital files. We may post blog entries accessible to anyone, or maintain friendships through online social networks. Each of these transactions leaves digital traces that can be compiled into comprehensive pictures of both individual and group behavior, with the potential to transform our understanding of our lives, organizations, and societies.

The capacity to collect and analyze massive amounts of data has transformed such fields as biology and physics. But the emergence of a data-driven "computational social science" has been much slower. Leading journals in economics, sociology, and political science show little evidence of this field. But computational social science is occurring—in Internet companies such as Google and Yahoo, and in government agencies such as the U.S. National Security Agency. Computational social science could become the exclusive domain of private companies and government agencies. Alternatively, there might emerge a privileged set of academic researchers presiding over private data from which they produce papers that cannot be critiqued or replicated. Neither scenario will serve the long-term public interest of accumulating, verifying, and disseminating knowledge.

What value might a computational social science—based in an open academic environment—offer society, by enhancing understanding of individuals and collectives? What are the



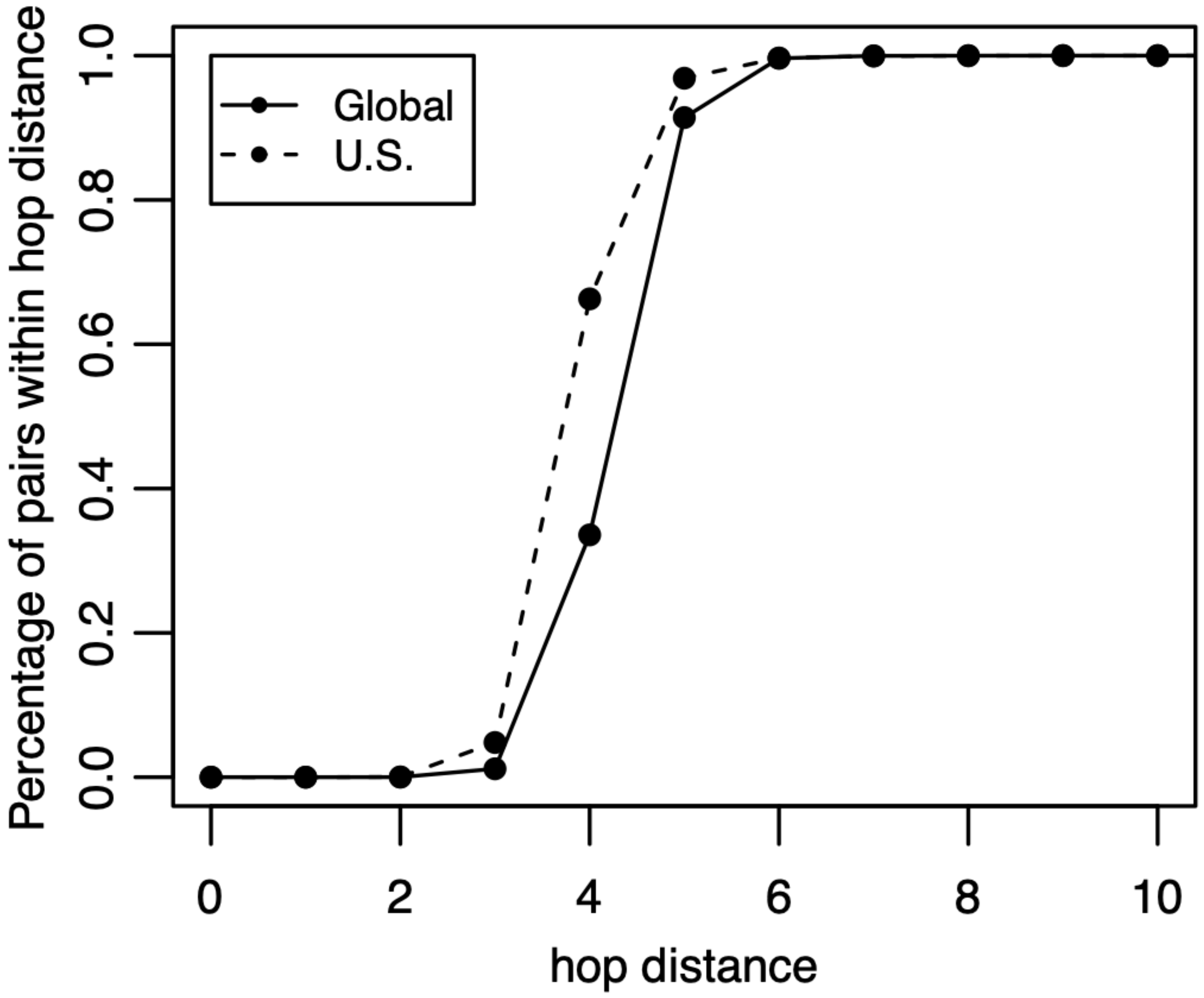
[1]Harvard University, Cambridge, MA, USA. [2]Massachusetts Institute of Technology, Cambridge, MA, USA. [3]University

Example: Facebook social network

- Facebook in 2011: 721 million users, 69 billion friendship connections
- “Six degrees of separation”

  → 5.2 in Milgram’s experiments in US in 50s

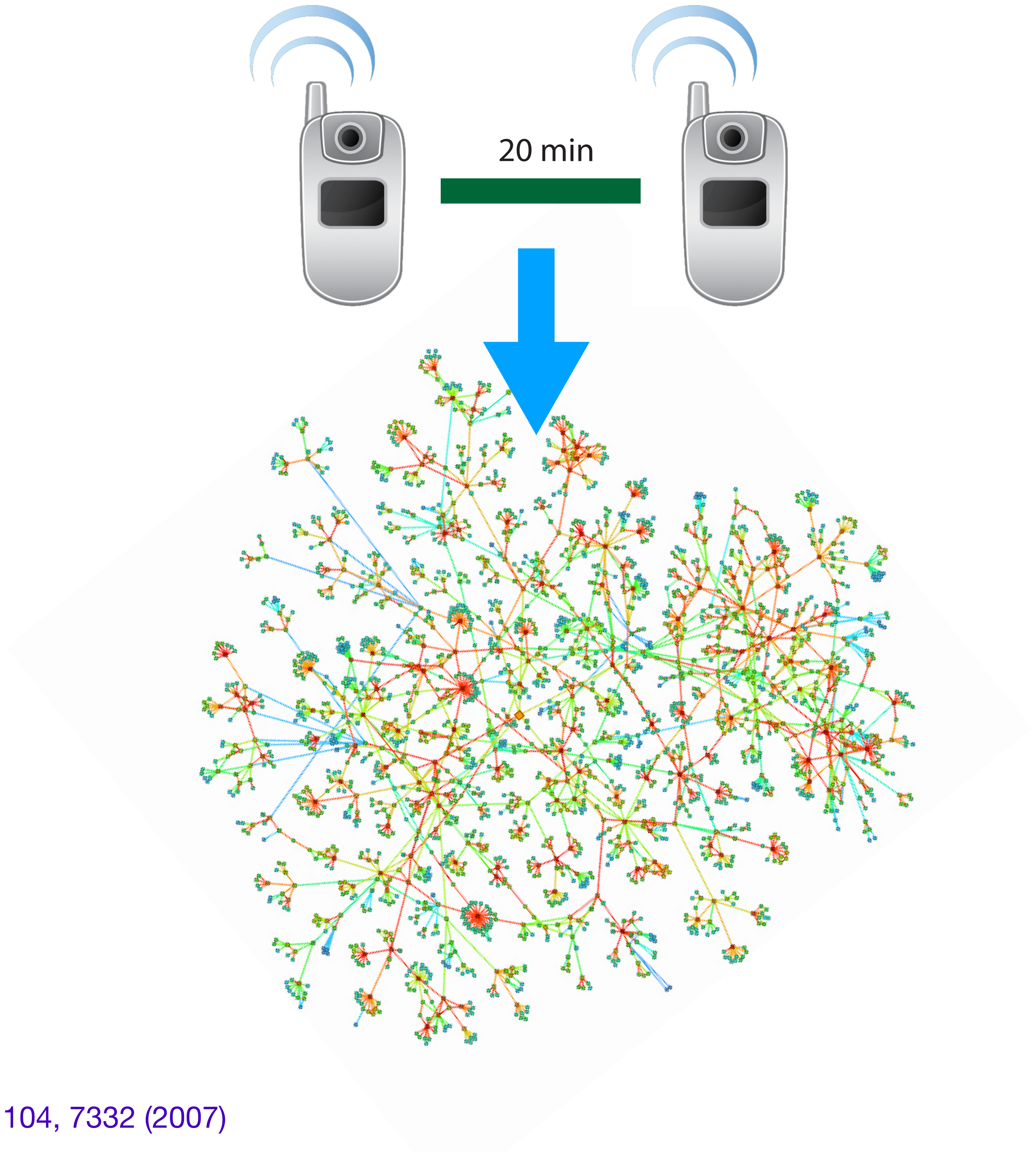
  → 3.74 in Facebook (in 2011)



J. Ugander, B. Karrer, L. Backstrom, C. Marlow "The Anatomy of the Facebook Social Graph”, https://arxiv.org/abs/1111.4503 (2011)

Example: Call data records

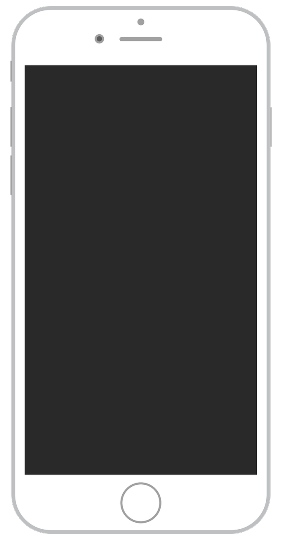
- Social network built out of data on call durations of customers of a mobille phone operator, 7 million customers, 18 months, 300 million calls
- “Granovetter’s hypothesis” in 1973: strong ties within social groups, weak ties between
- Mobile data proved the hypothesis at a scale of a country



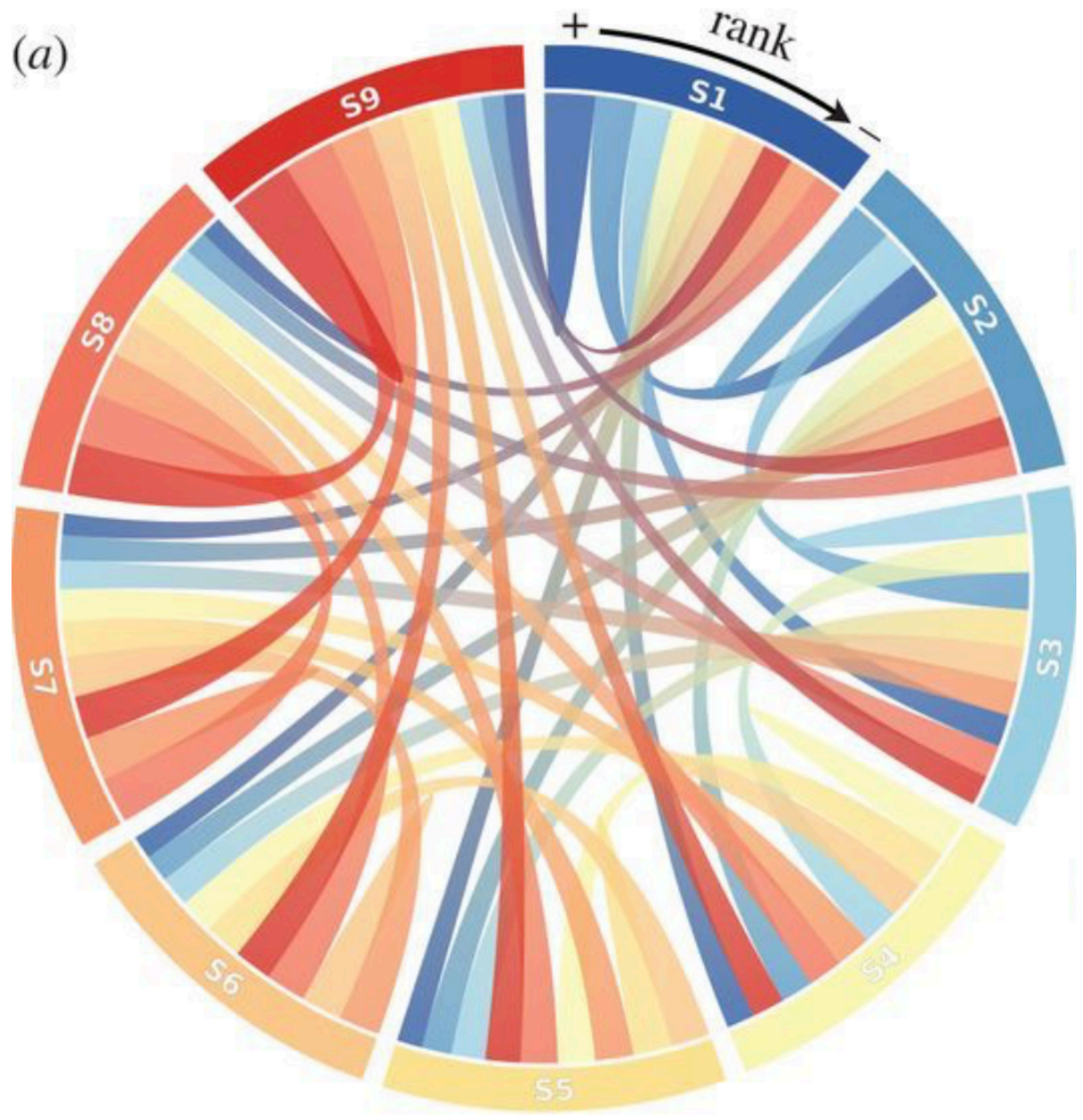
Onnela, et al., Proc. Natl. Acad. Sci. (USA) 104, 7332 (2007)

Example: phone calls + credit data

- ~8 billion calls and sms between 112 mobile phone customers in Mexico
- Combined with banking data (purchases, loans etc) of 6 million customers



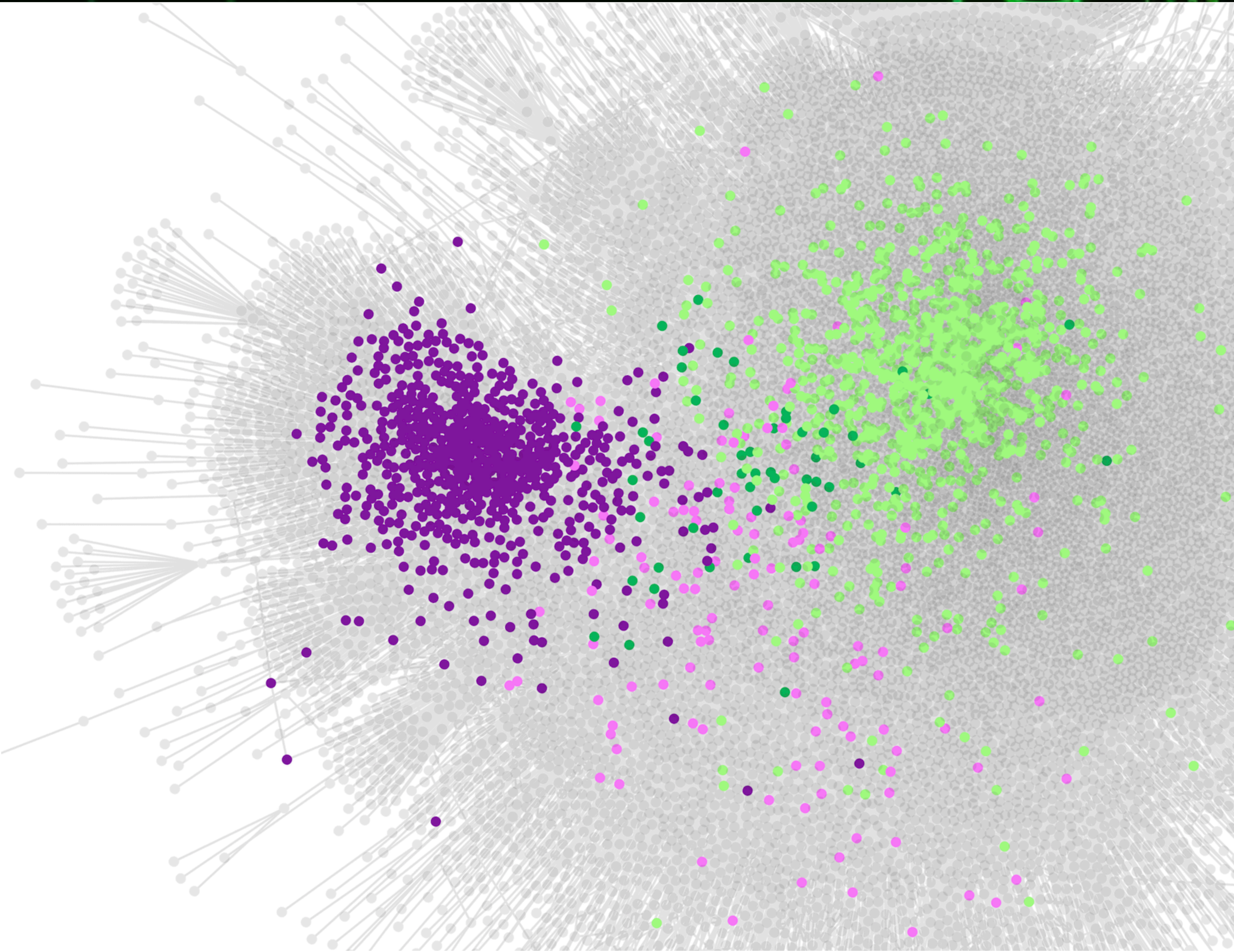
Map of socioeconomic stratification:



Leo, Y., Fleury, E., Alvarez-Hamelin, J. I., Sarraute, C., & Karsai, M. (2016). Socioeconomic correlations and stratification in social-communication networks. *Journal of The Royal Society Interface*, *13*(125), 20160598.

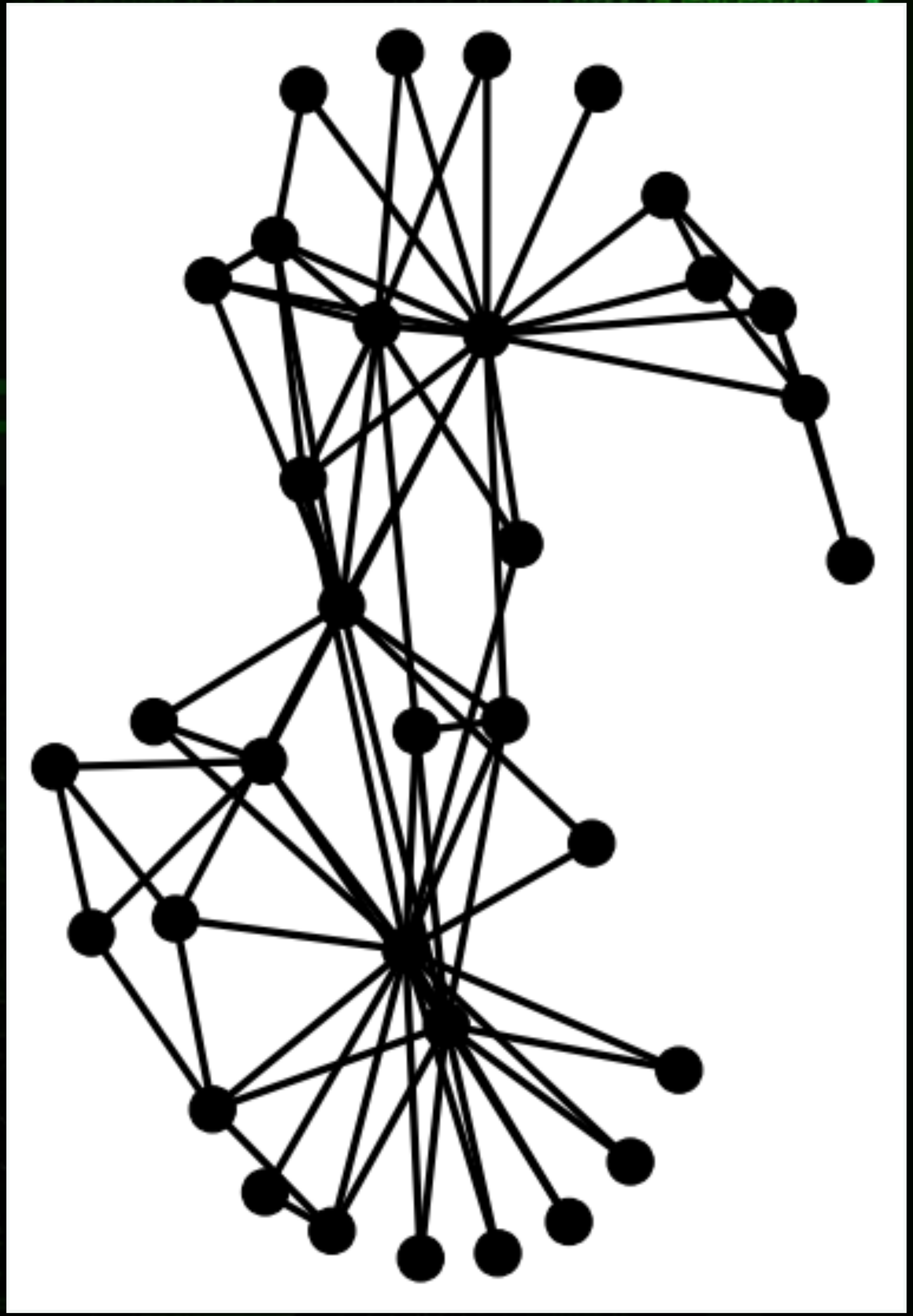
Example: Twitter data

- Twitter has around 200 million daily users
- Figure: political discussion network in Finland during the 2019 elections



THY Chen, A Salloum, A Gronow, T Ylä-Anttila, M Kivelä "Polarization of Climate Politics Results from Partisan Sorting: Evidence from Finnish Twittersphere." Global Environmental Change 71 (2021): 102348.

How to analyse social networks?



History: Sociograms

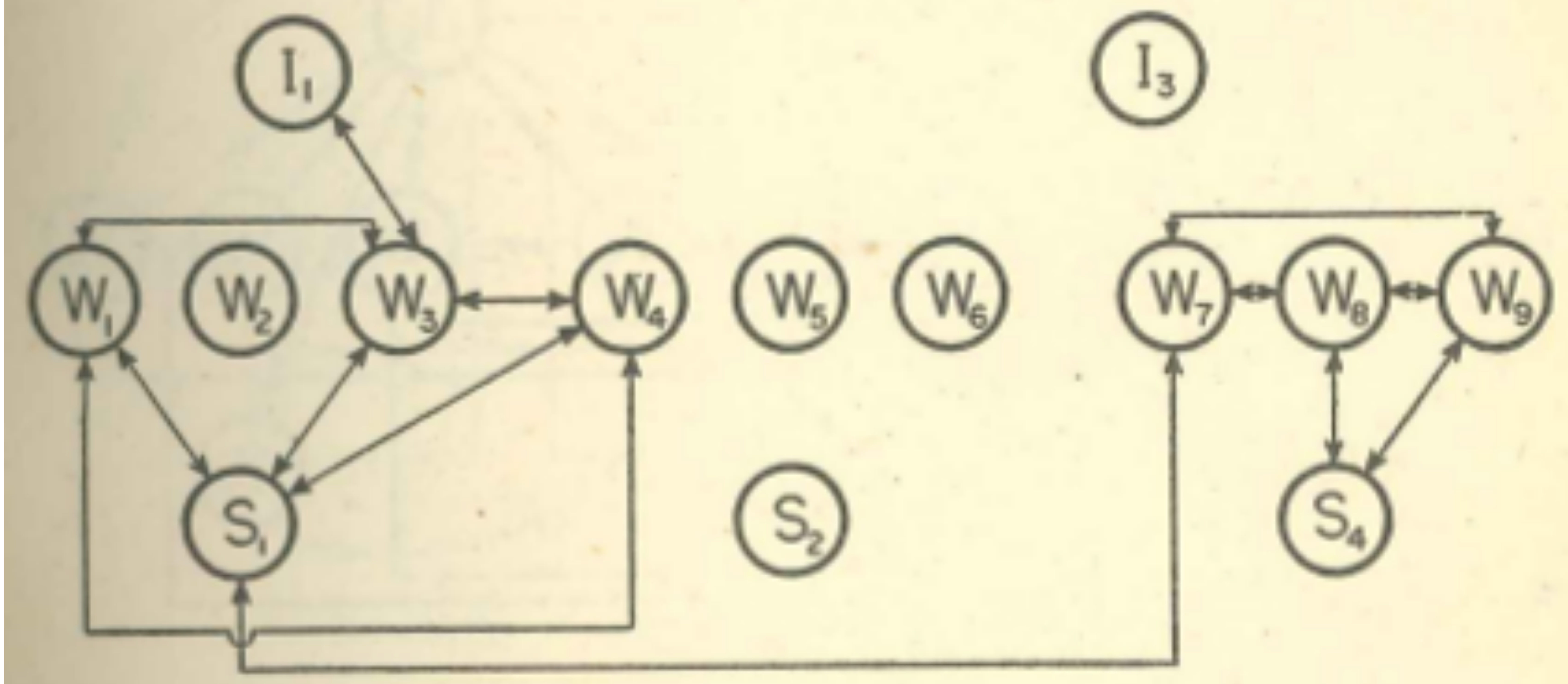


Fig. 6. Bank Wiring Room: Men who were friends

G. C. Homans. "Human Group", Routledge 1951

F. Roethlisberger, W. Dickson. "Management and the worker", Cambridge University Press 1939

What kind of structure is there in social networks?

Why is the structure important?

Plan for this lecture:

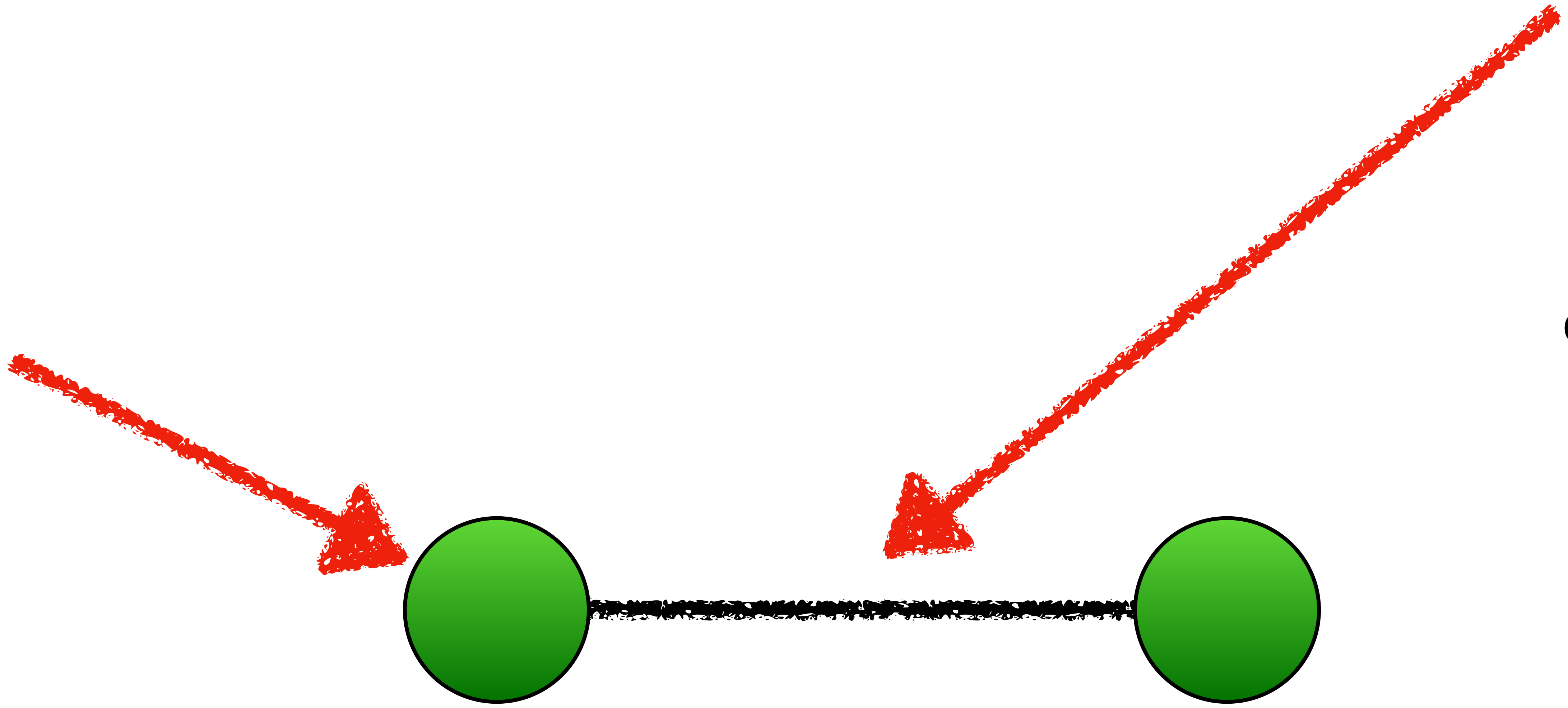
From small-scale structure to large-scale structure

Basic building blocks

Node

=

person
user account
organisation
…



Link

=

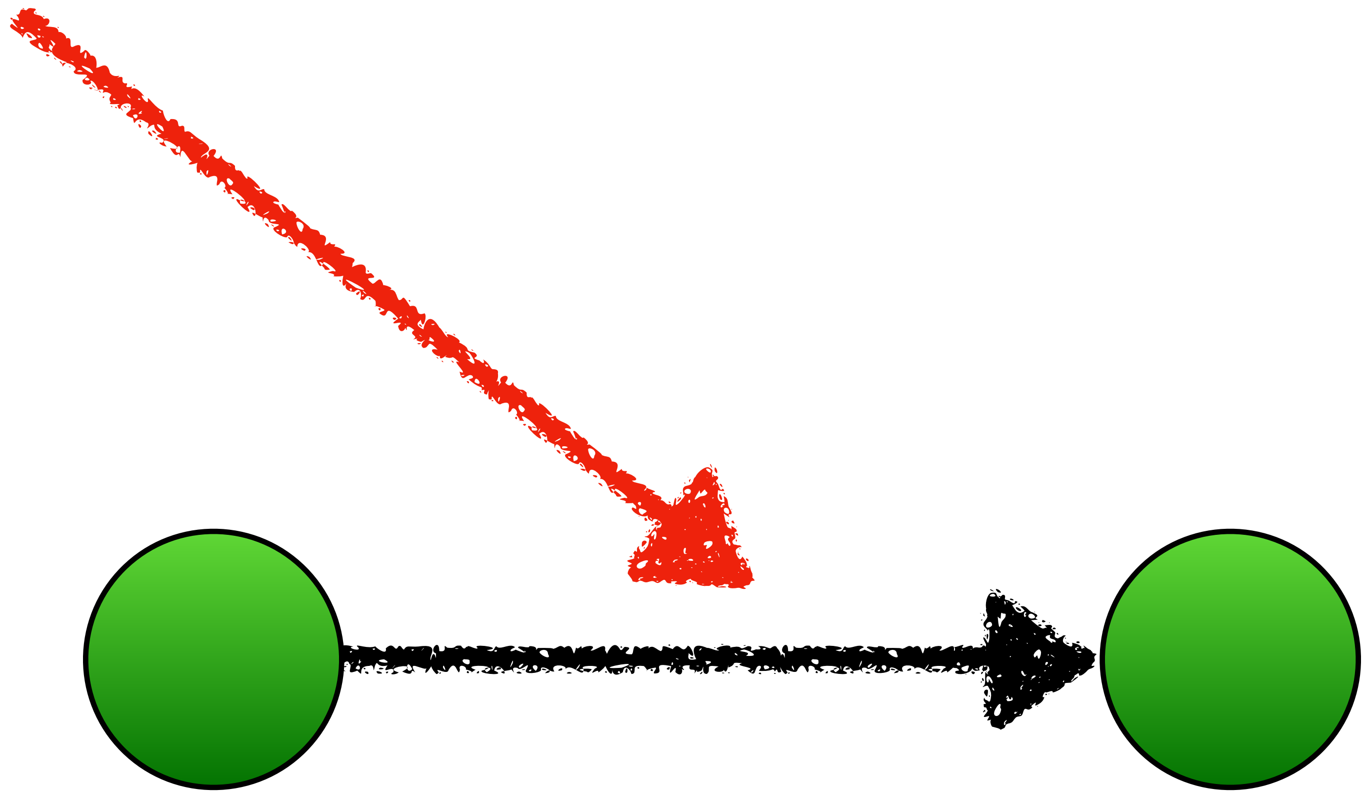
interaction
relationship
friendship
professional contact
messaged each other
online platform connection
…

Basic building blocks

Directed link

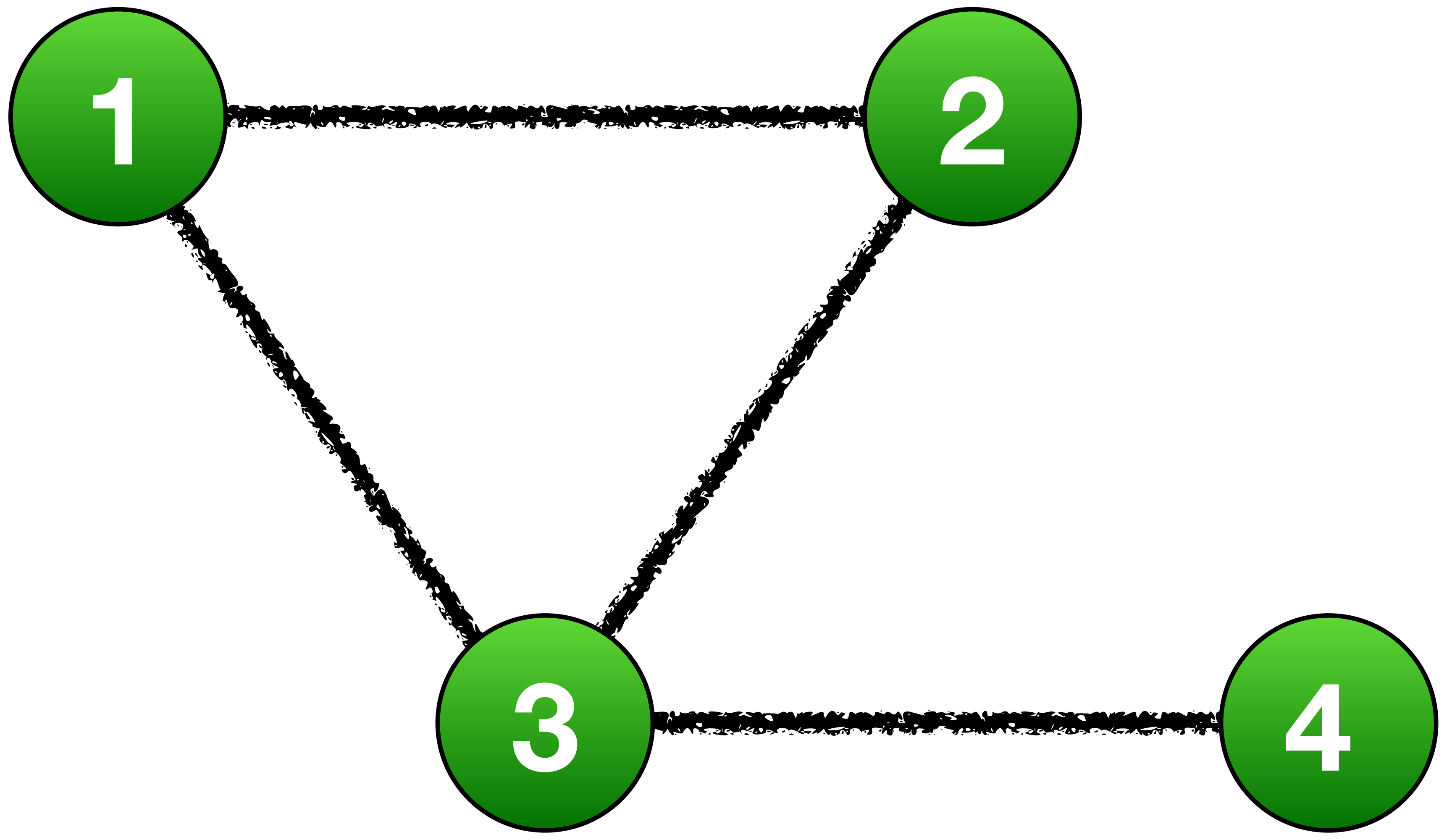
=

following relationship
message in one direction
asks advice from
gets information from
...



Network = graph

- Graph G = (V,E)
- V = set of nodes/vertices
- E = set of links/edges/ties (directed or undirected)

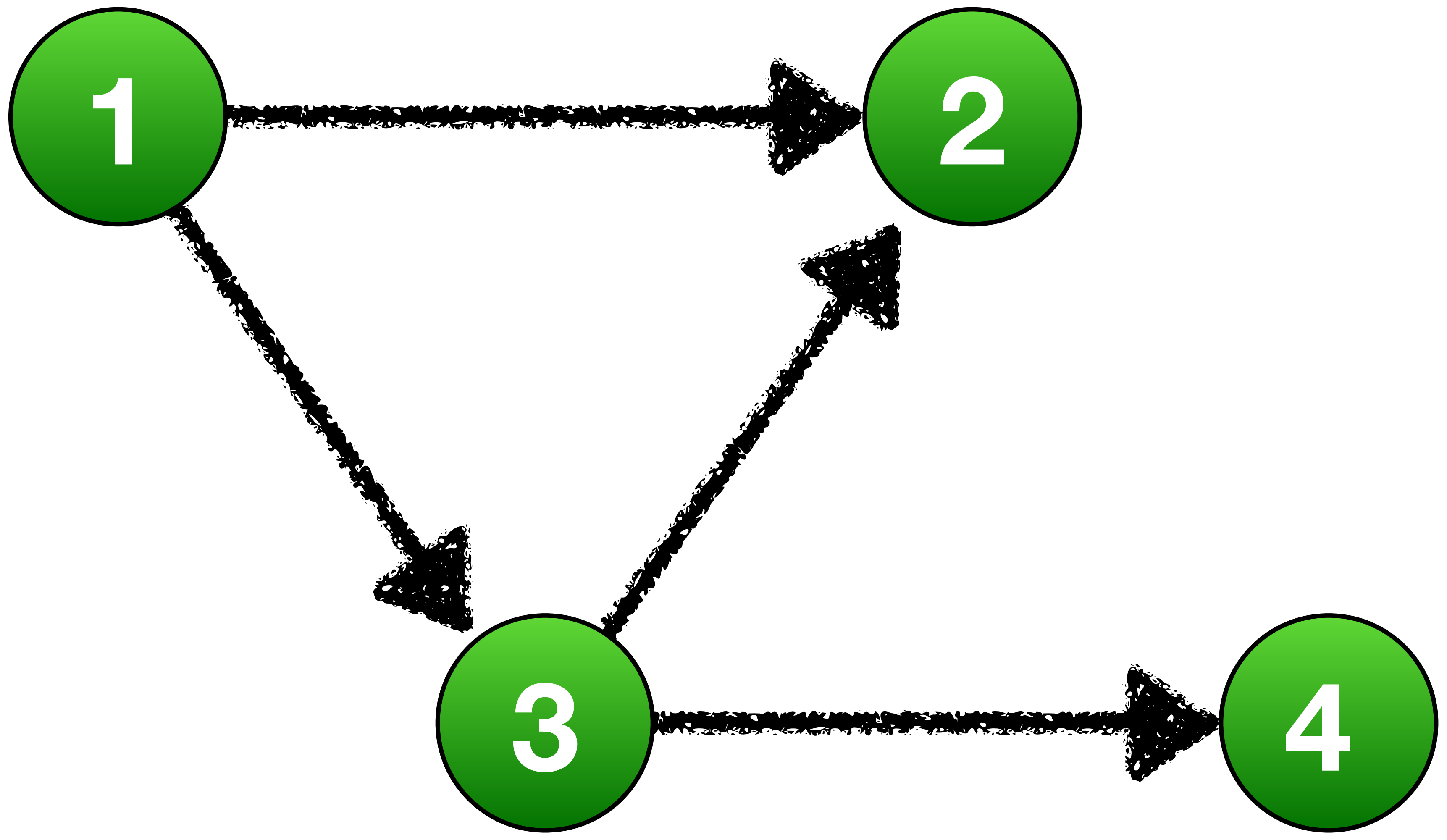


V = {1, 2, 3, 4}

E = {(1,2), (1,3), (2,3), (3,4)}

Network = graph

- Graph G = (V,E)
- V = set of nodes/vertices
- E = set of links/edges/ties (directed or undirected)

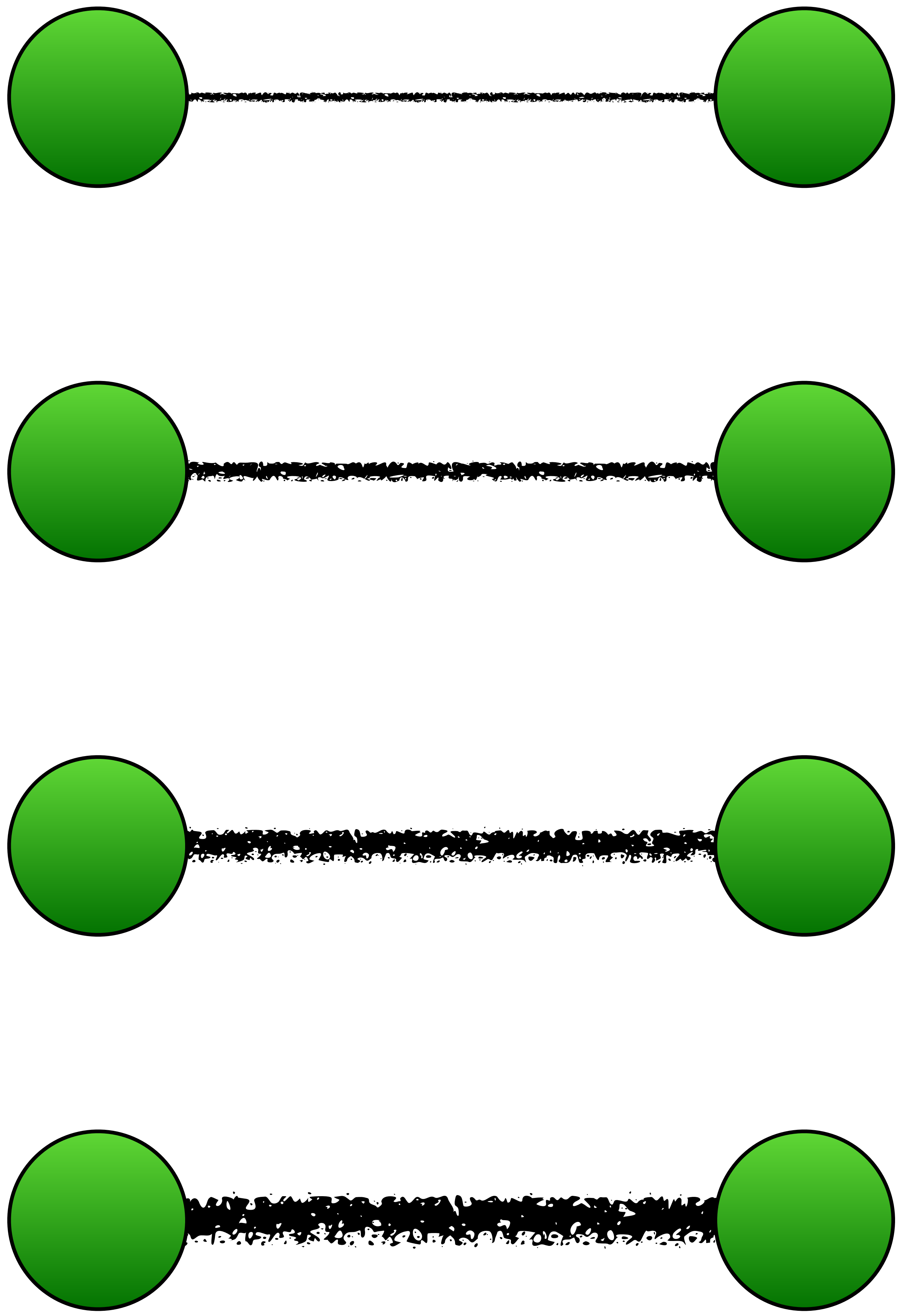


V = {1, 2, 3, 4}

E = {(1→2), (1→3), (3→2), (3→4)}

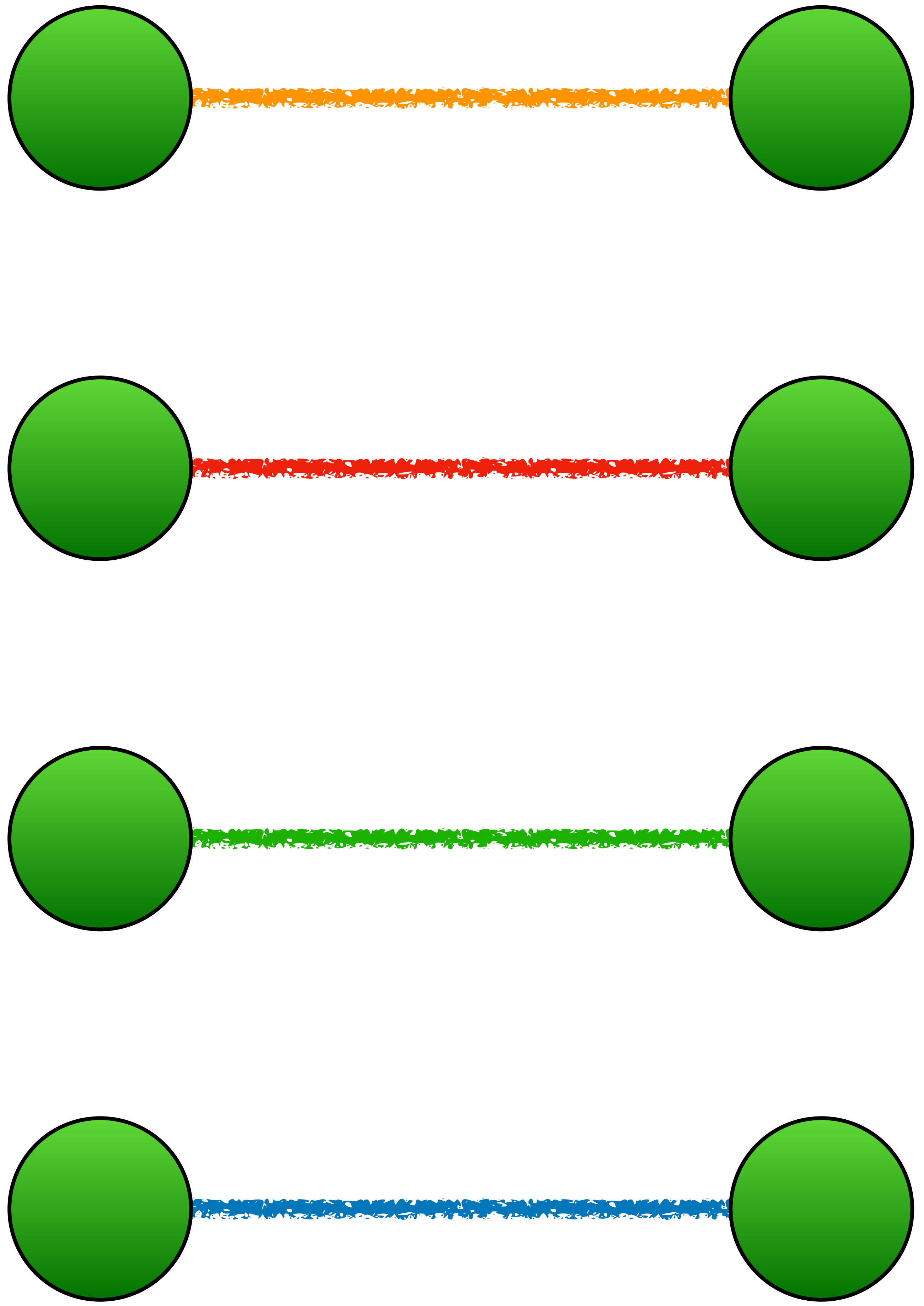
Tie strengths

- Not all social connections are equally strong
- “strength of a tie”:
  - how emotionally close
  - how much time you spend
  - how many interactions
  - …
- Typically represented as numerical value in data: *weight of the link*



Type of a tie

- Connections/interactions have different meanings & contexts
- Examples:
  - Types of relationships: frienship, professional, family, …
  - Content of communication: politics, sports, science, …
  - Actions: ask advice, plays games with, argues with, …
- Formally: Multiple types of links → multiplex networks



M. Kivelä, A. Arenas, M. Barthelemy, J.P. Gleeson, Y. Moreno, M.A. Porter: *Multilayer networks,* J. Complex Netw. 2(3): 203-271 (2014)

Type of a tie

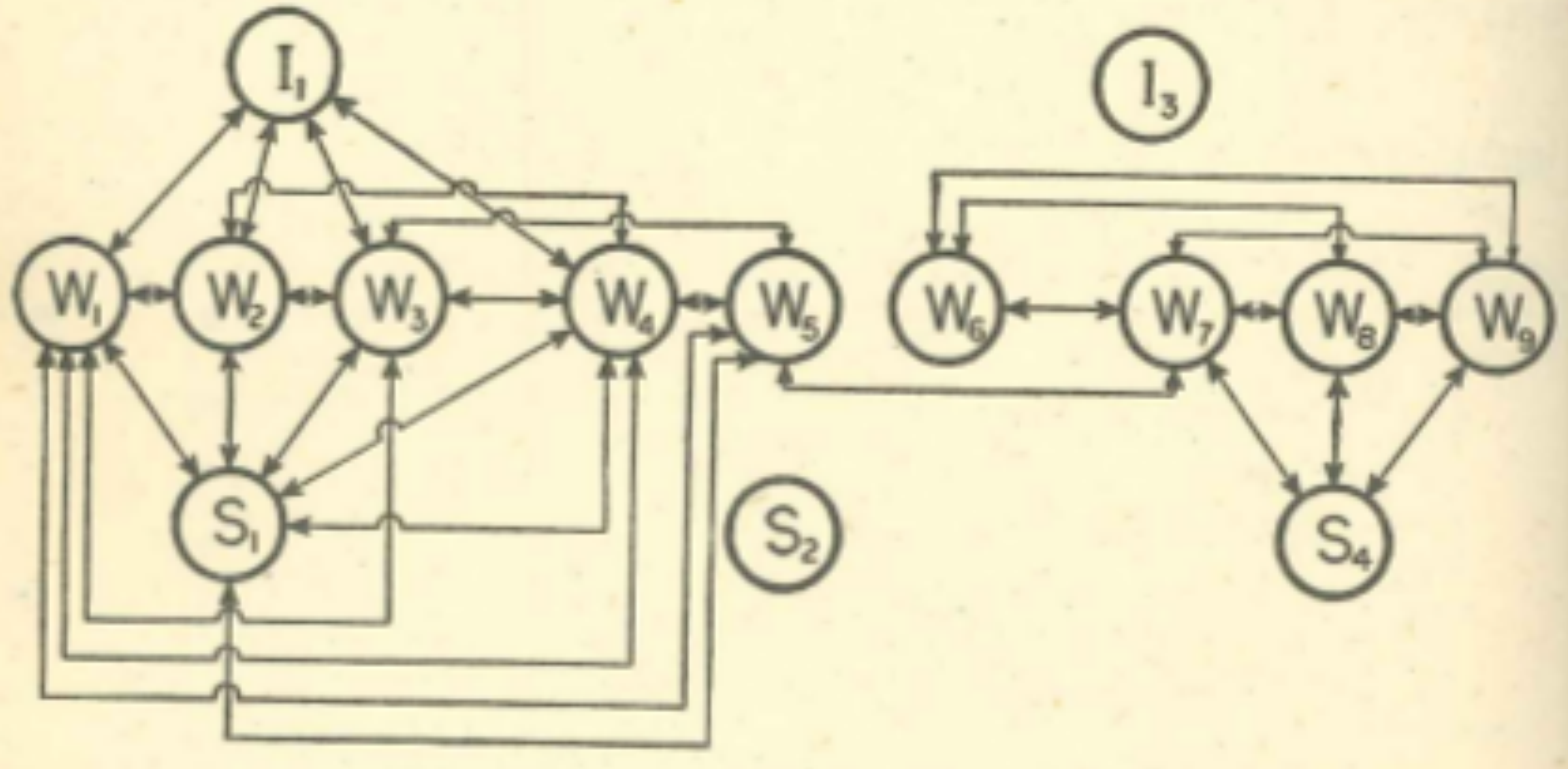


Fig. 5. Bank Wiring Room: Men who played games together

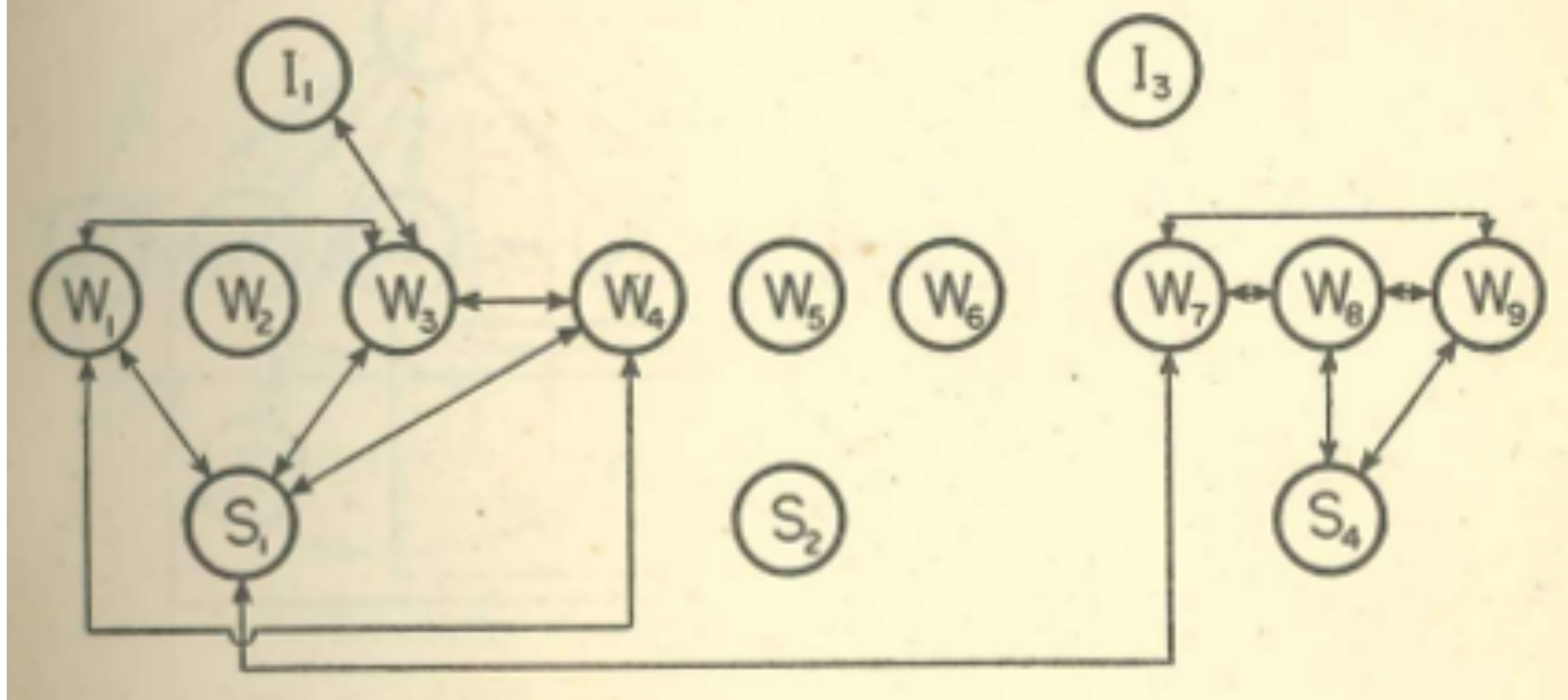


Fig. 6. Bank Wiring Room: Men who were friends

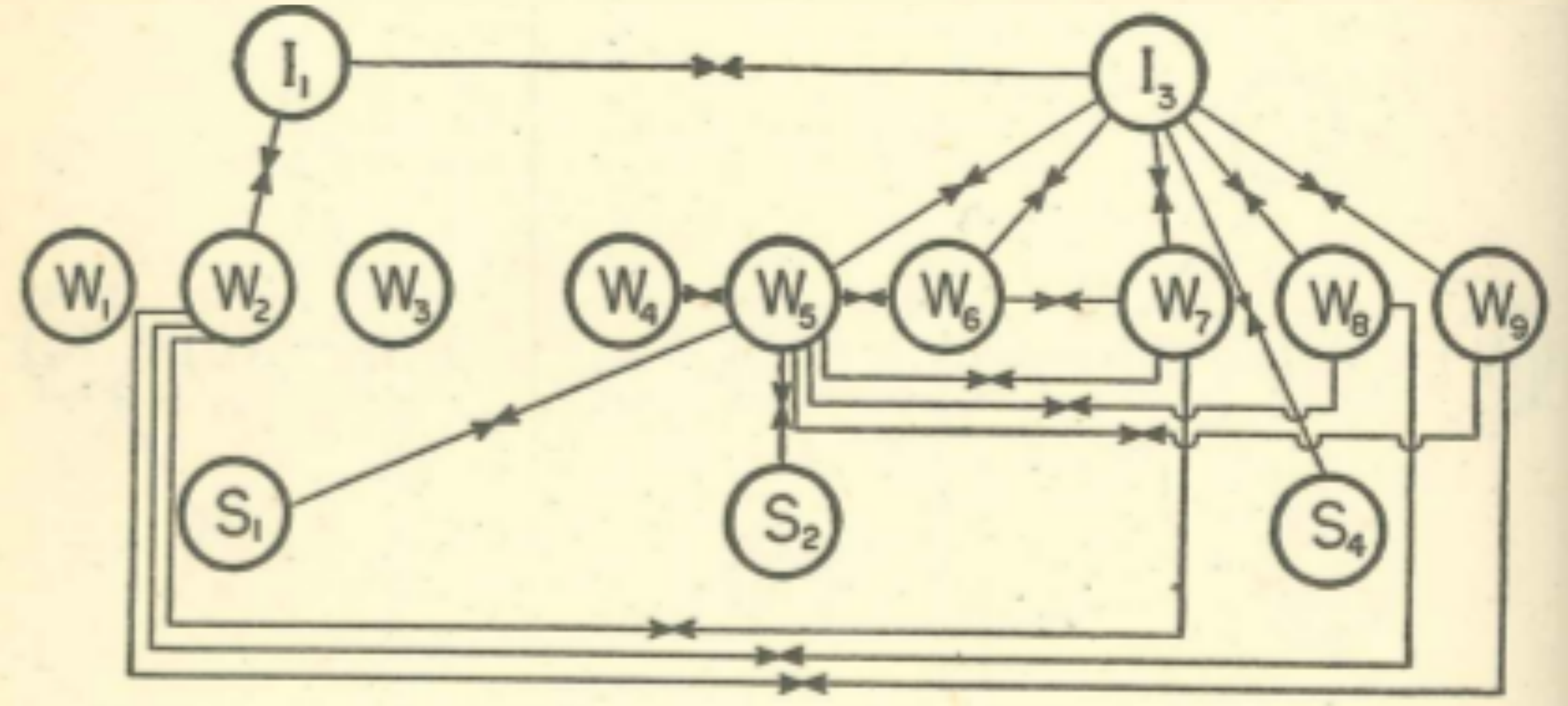


Fig. 7. Bank Wiring Room: Men who were antagonistic to one another

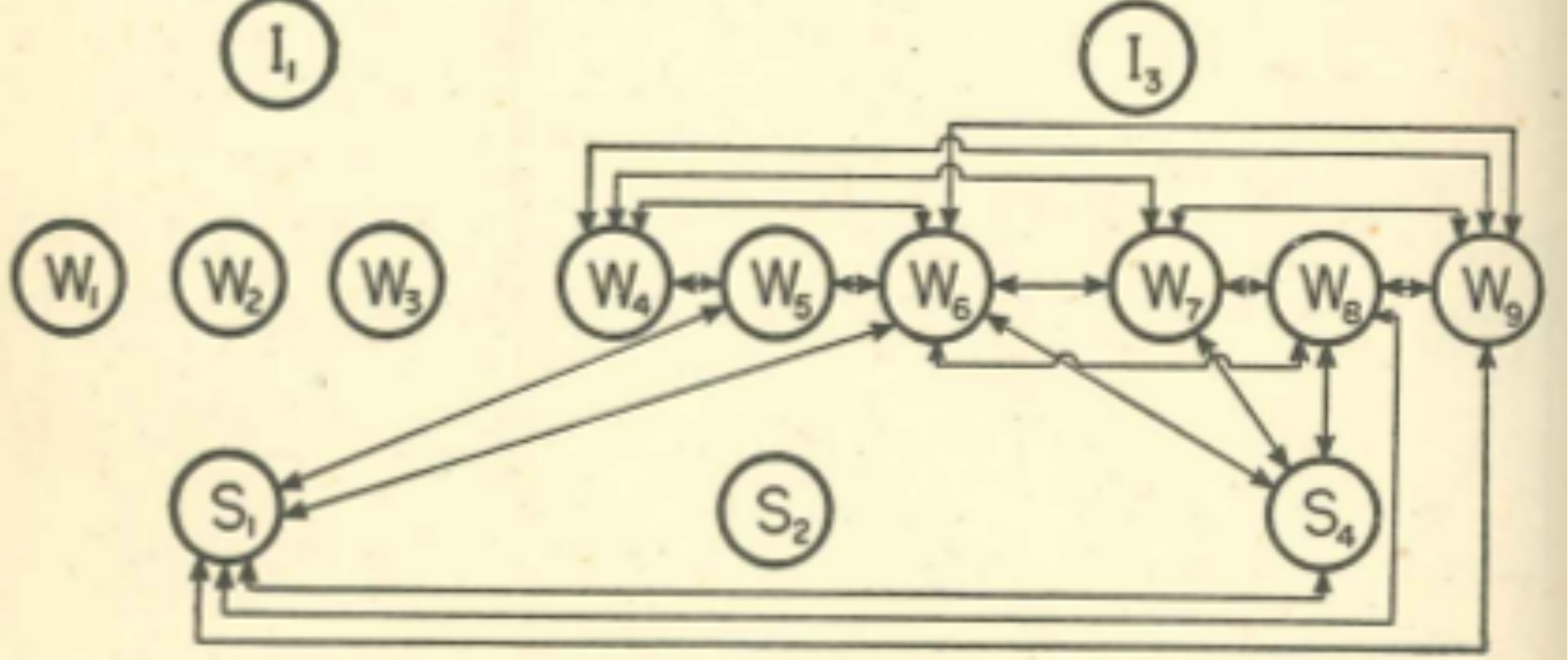
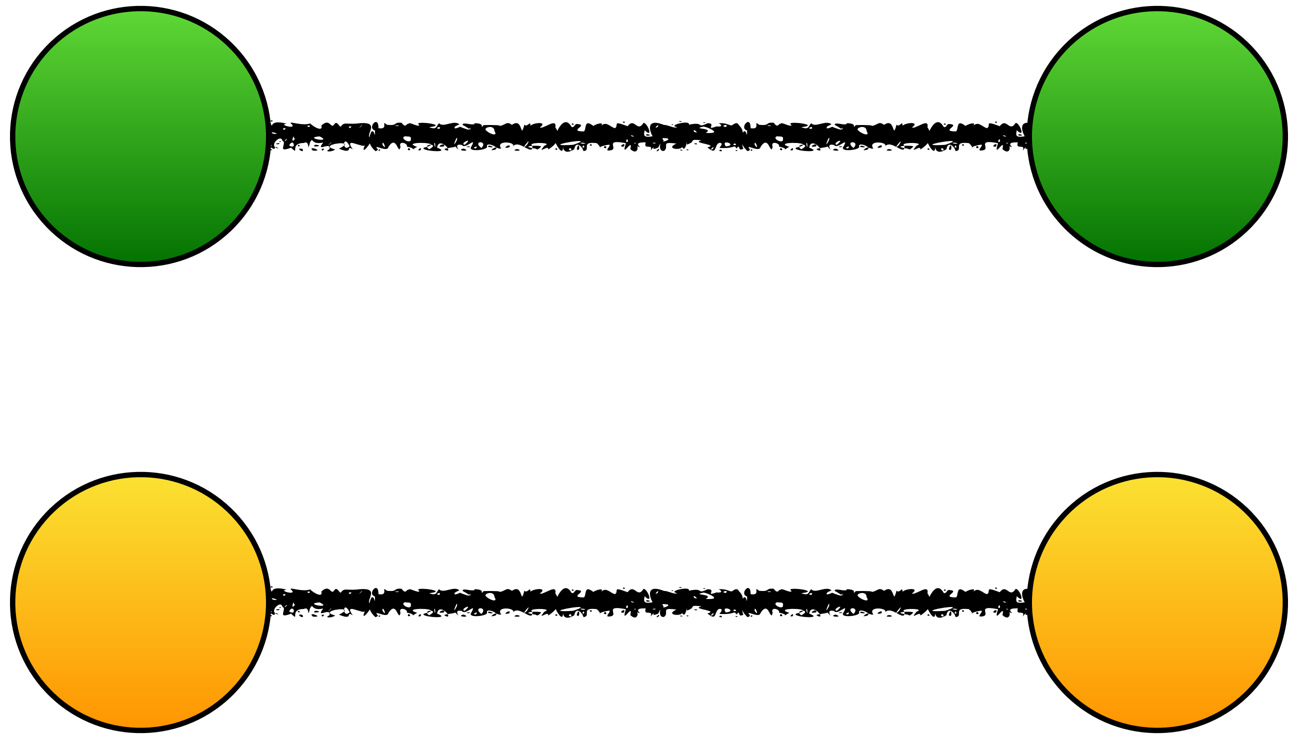


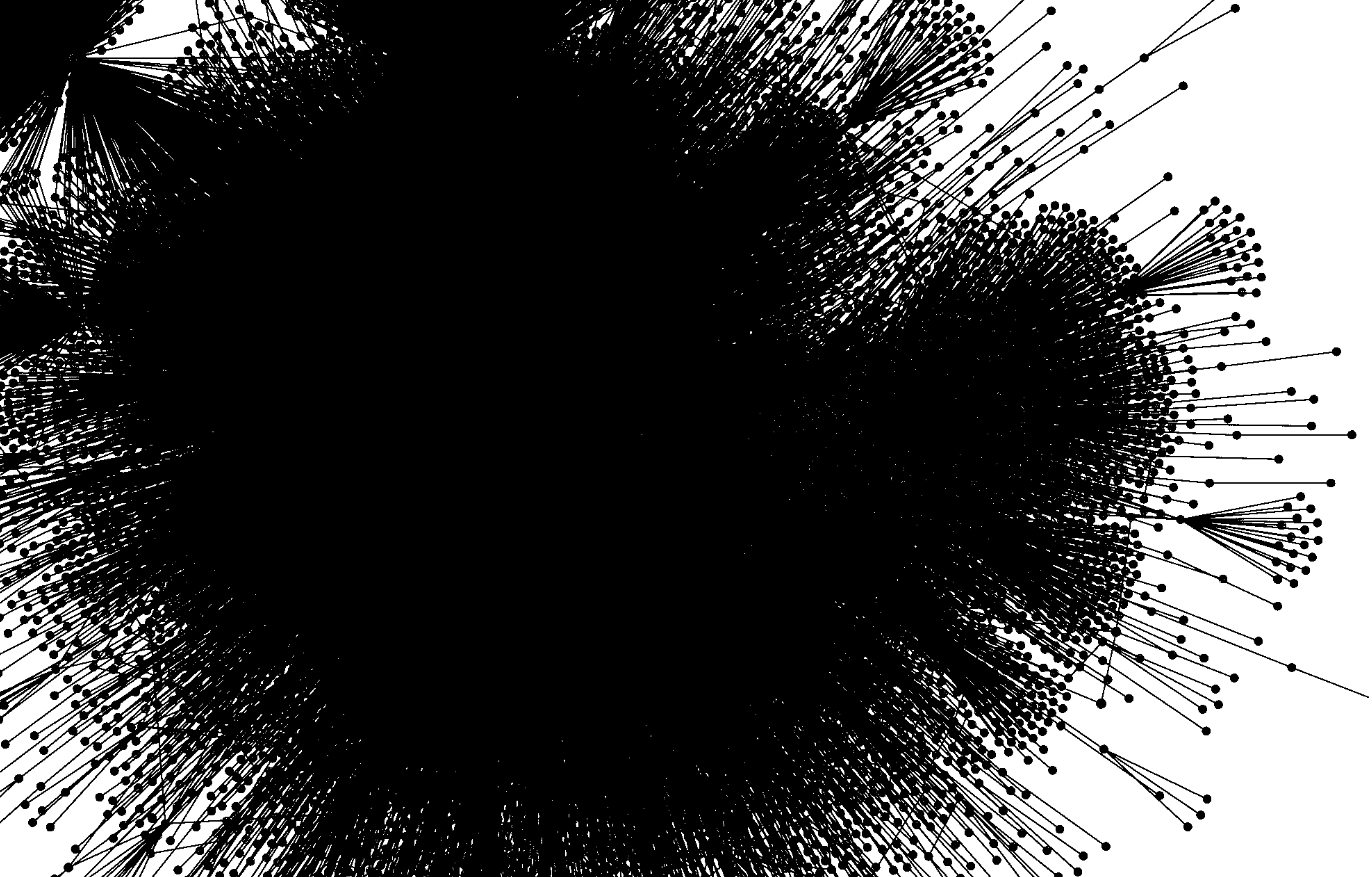
Fig. 9. Bank Wiring Room: Men who got into arguments about windows

Homophily

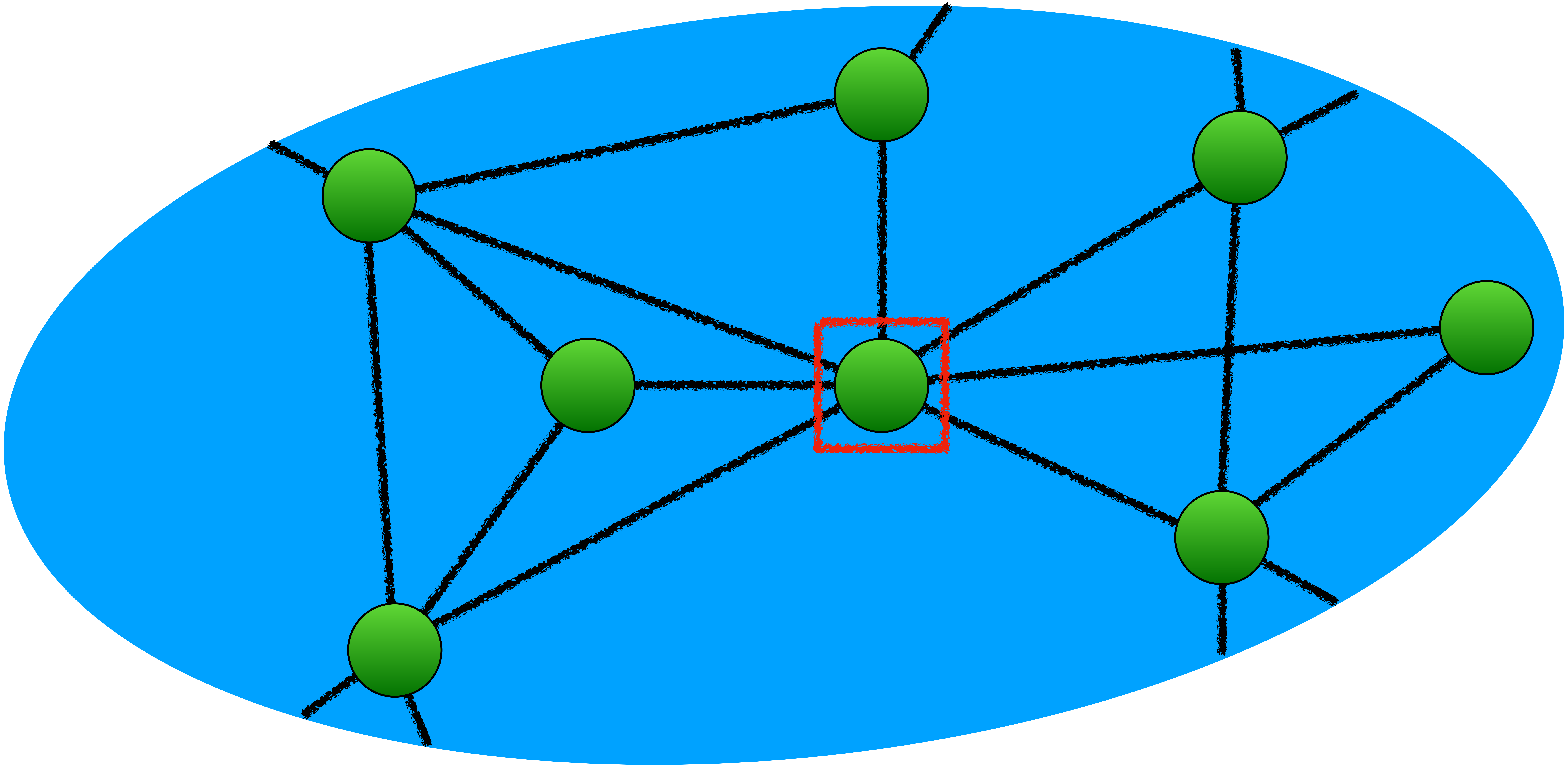
- Similar people tend to be connected to each other
- “Birds of the feather flock together”
- Observed in most social networks
- Mechanism: choice homophily → lecture 9



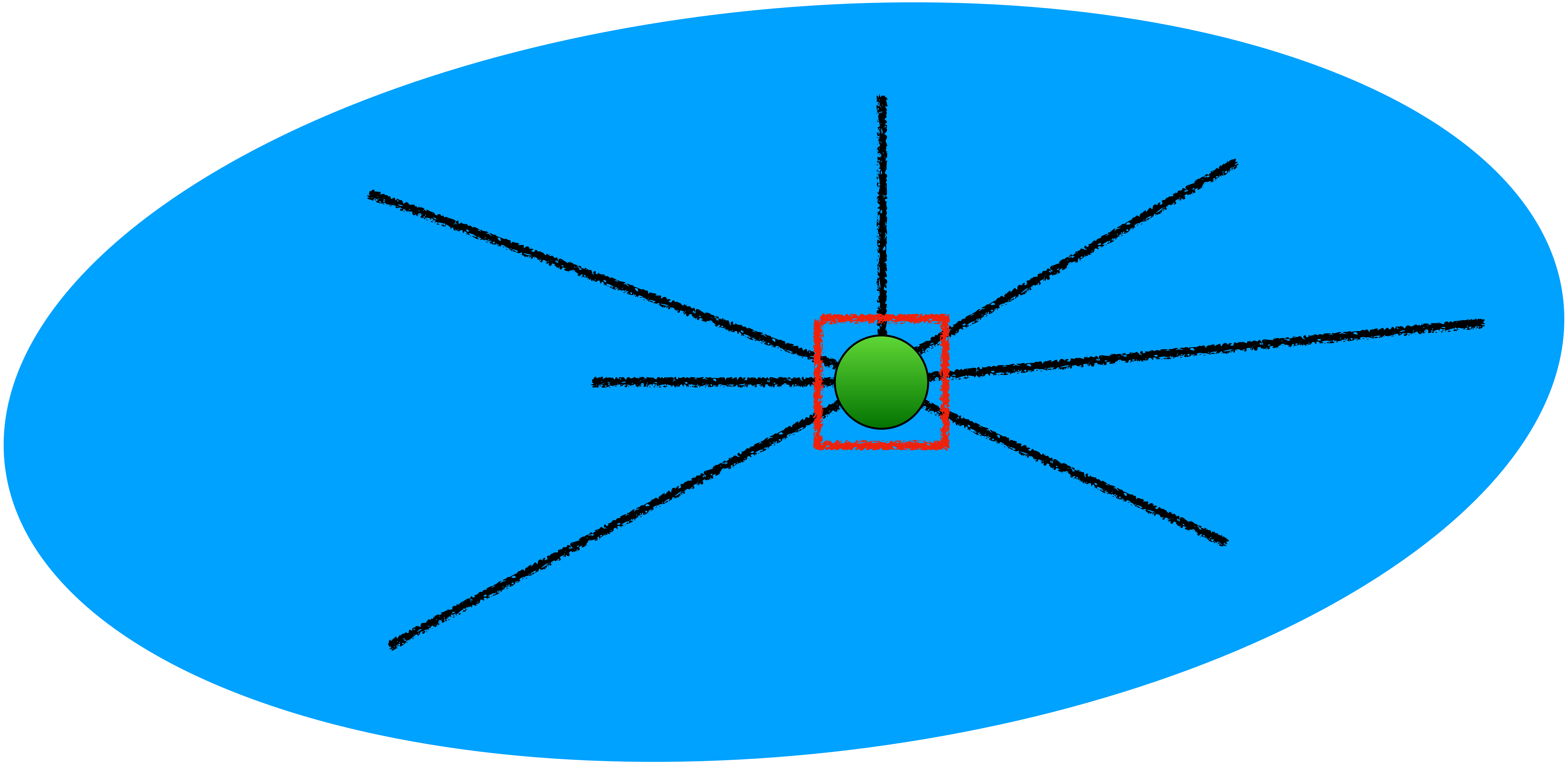
M McPherson, L Smith-Lovin, JM Cook "Birds of a Feather: Homophily in Social Networks”, Annual Review of Sociology Vol. 27:415-444 (2001)



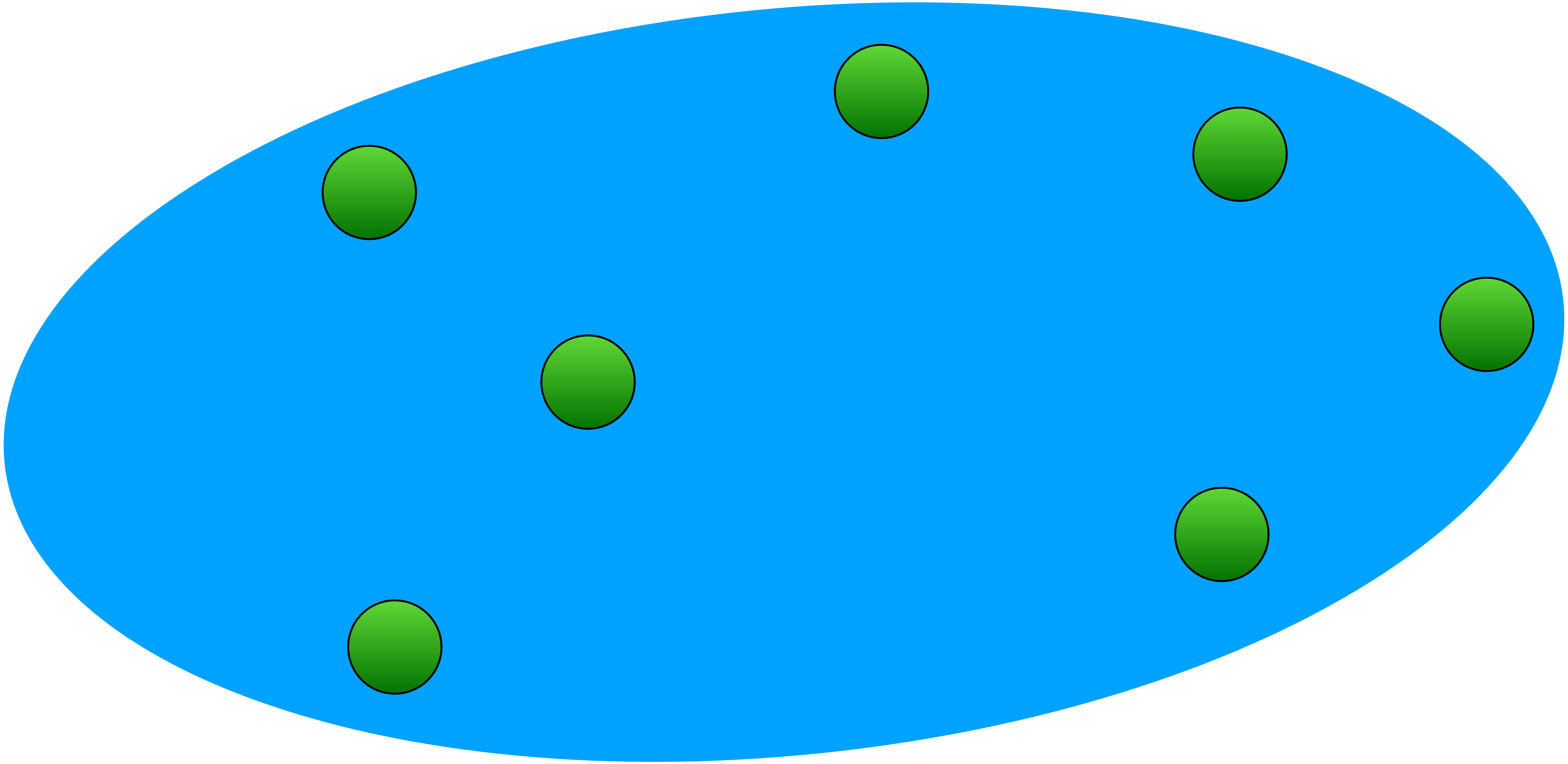
Neighborhood around a node



Degree = number of links

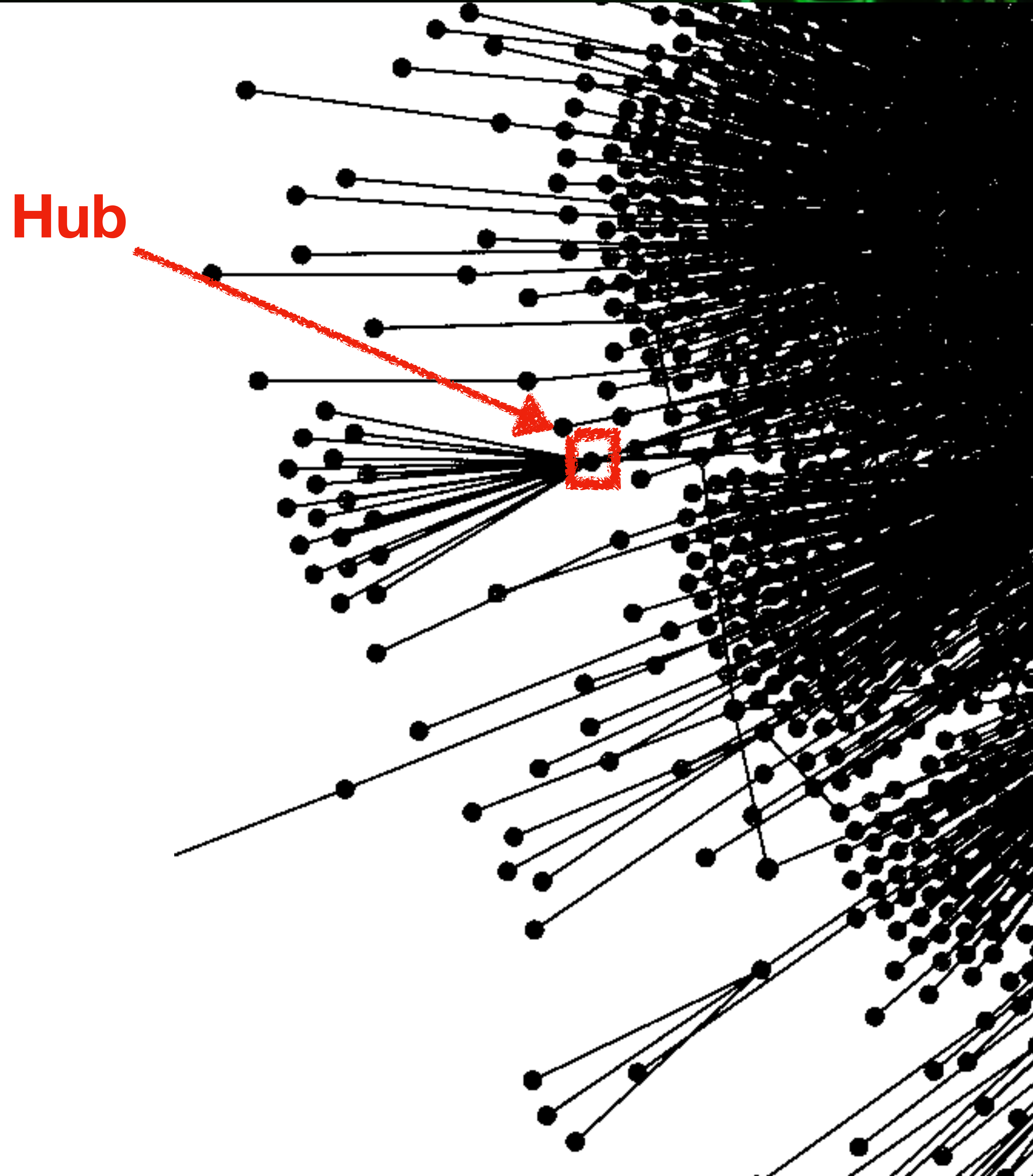


Degree = number of neighbours



Degrees & hubs

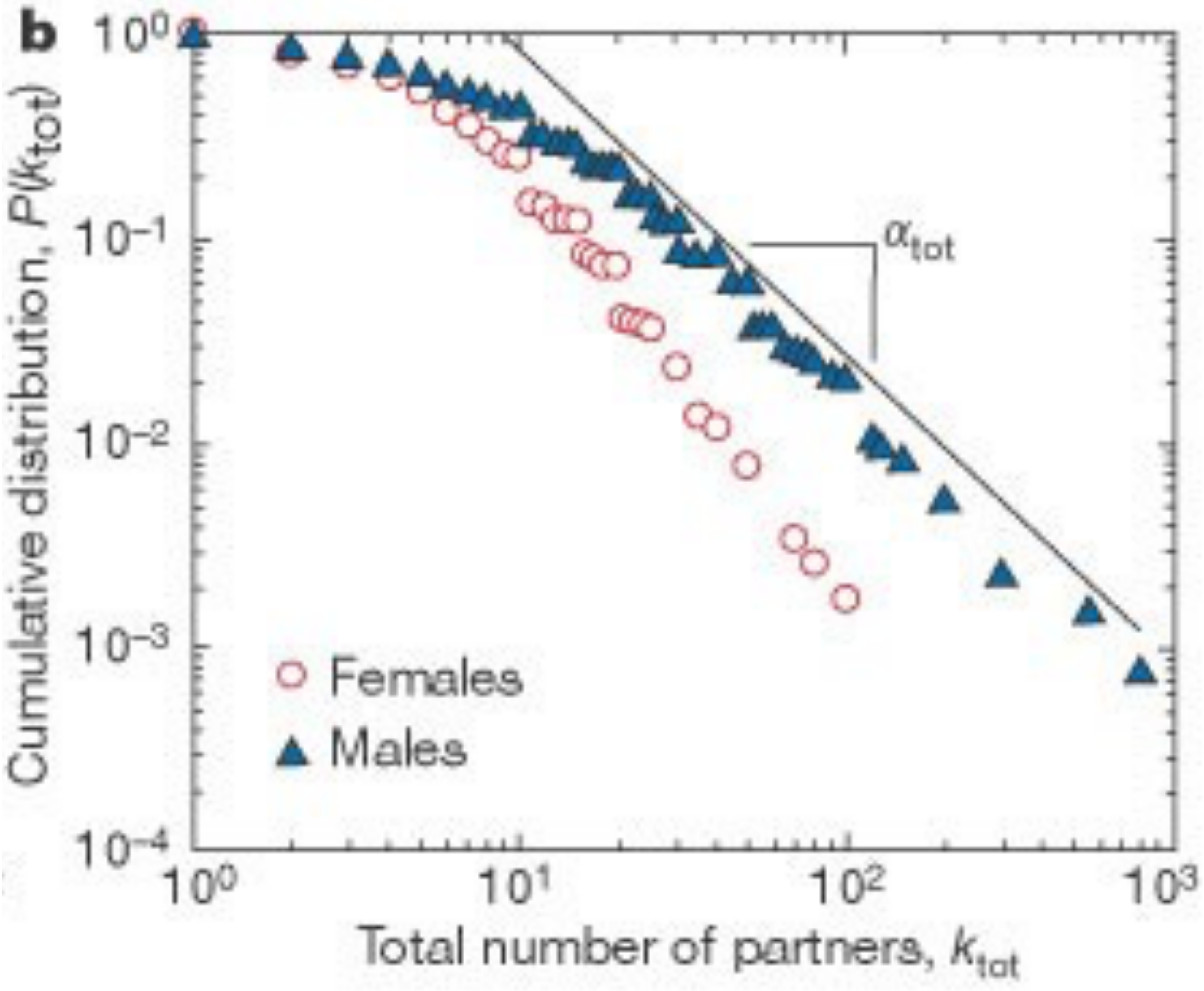
- Degree = k_i = number of connections/neighbors node i has
- Hub: a node that has very high degree
- Typically most nodes have very low degree, few hubs



Scale-free networks?

- Scale-free networks: power-law degree distribution: $P(k) \sim k^{-\alpha}$
- Often plotted in double logarithmic axis → straight line: $\log P(k) \sim -\alpha \log k$
- Many networks reported as scale-free in the literature

→ Typically in reality a “scale-free tail” (no power-law for low k values)

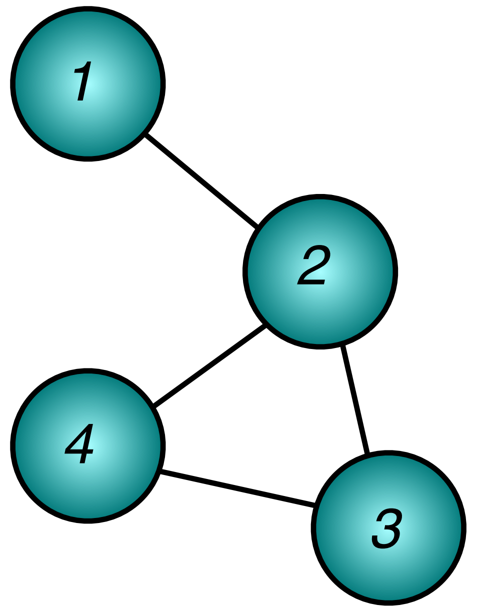


F. Liljeros et al., Nature 2001

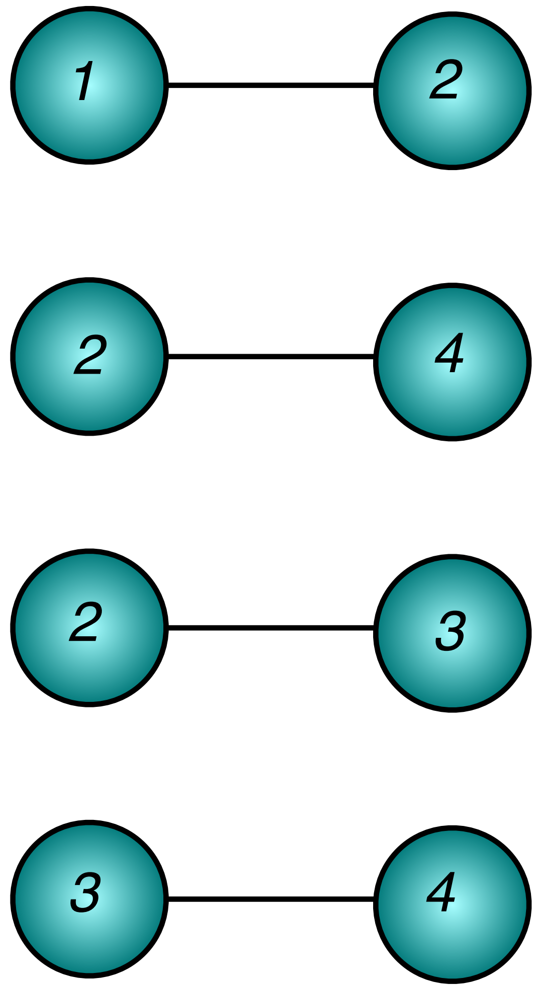
Friendship paradox

If you follow a random link, what is the probability that node i with degree k_i is picked?

Example network:



Each link picked with probability 1/4



After that, each node is picked with probability 1/2

$$p_1 = \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{8}$$
$$p_2 = 3 \cdot \frac{1}{4} \cdot \frac{1}{2} = \frac{3}{8}$$
$$p_3 = 2 \cdot \frac{1}{4} \cdot \frac{1}{2} = \frac{2}{8}$$
$$p_4 = 2 \cdot \frac{1}{4} \cdot \frac{1}{2} = \frac{2}{8}$$

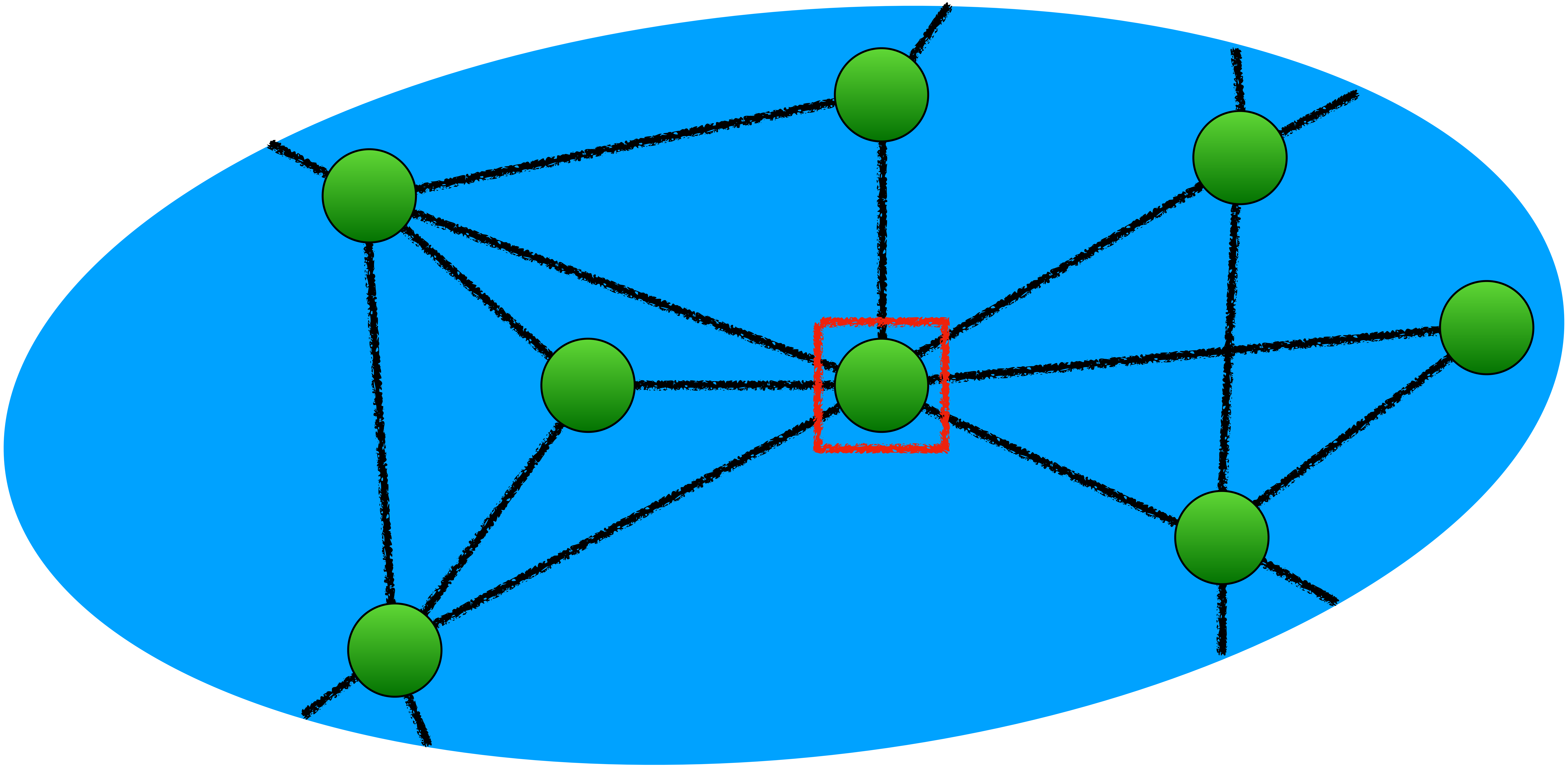
$$p(\text{'follow random link, reach node i'}) = \frac{k_i}{\sum_j k_j}$$

Friendship paradox: implications

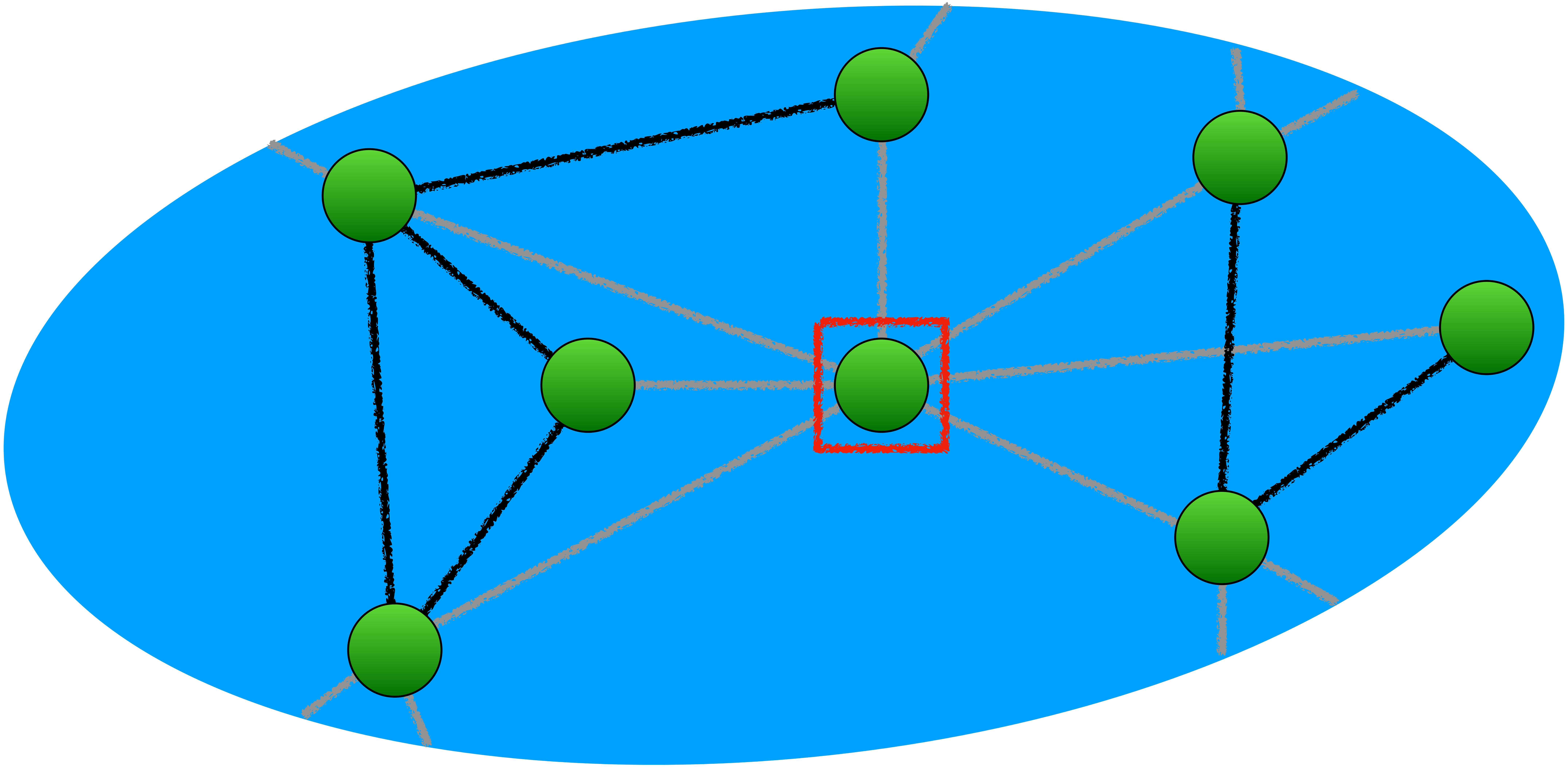
- Any process where links are followed is likely to lead to hubs
- **Spreading processes** lead to hubs (which are also effective spreaders)
- **Shortest paths** go through hubs
- **Meeting new people** through existing connections leads to new connections to hubs

→ More on lecture 9

Neighborhood around a node



How dense is the network around a node?



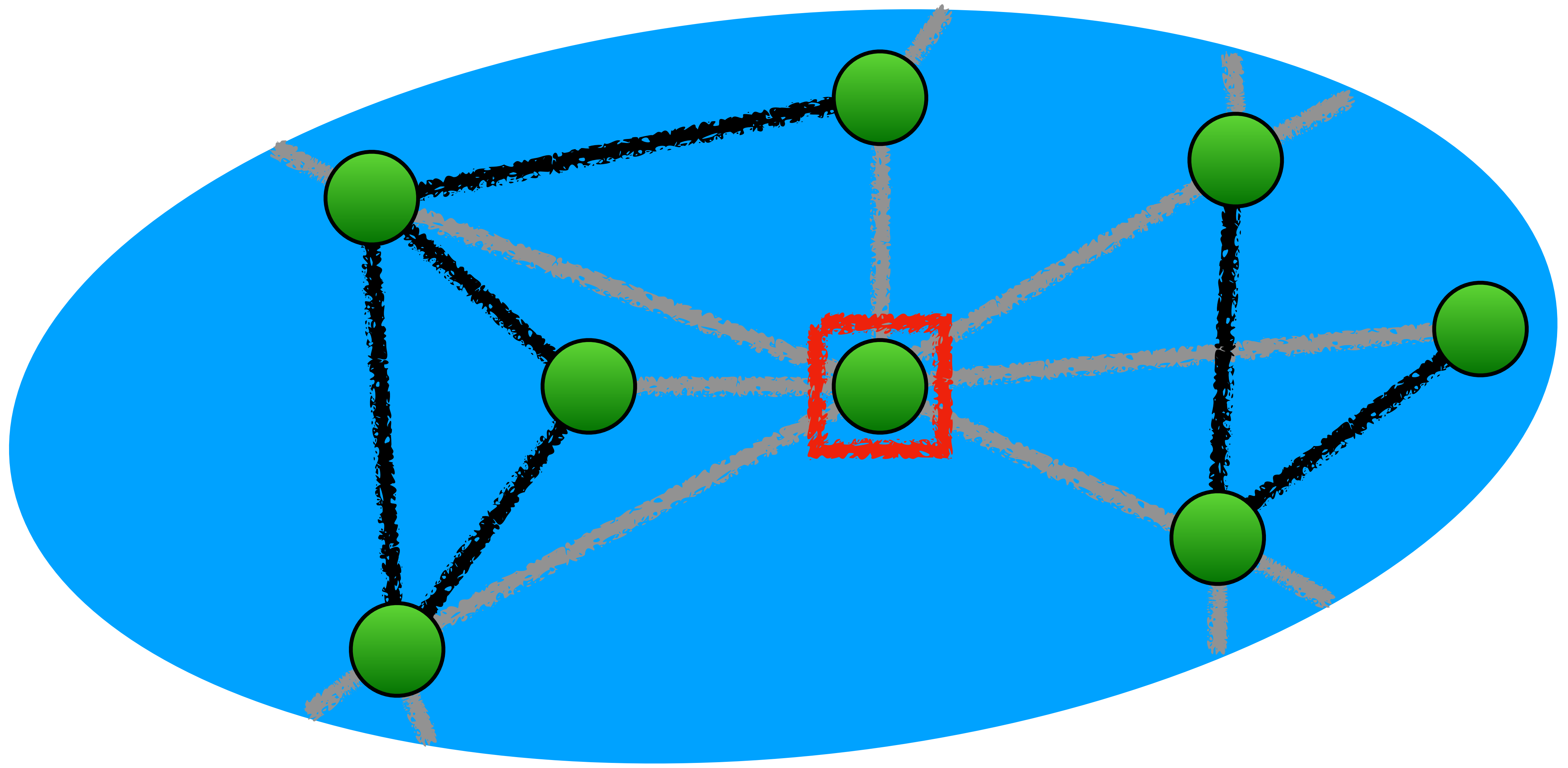
Clustering coefficient

- Clustering coefficient = density of the neighbourhood = # links between neighbouring nodes / # pairs of neighbours

- = # triangles / # of two-stars

- Social networks have high clustering coefficients:

  - Common friends, social groups, meeting people through existing connections…



$$c_i = \frac{E_i}{\binom{k_i}{2}}$$

$$E_i = 6 \qquad \binom{k_i}{2} = 7 \times 6/2 = 21$$

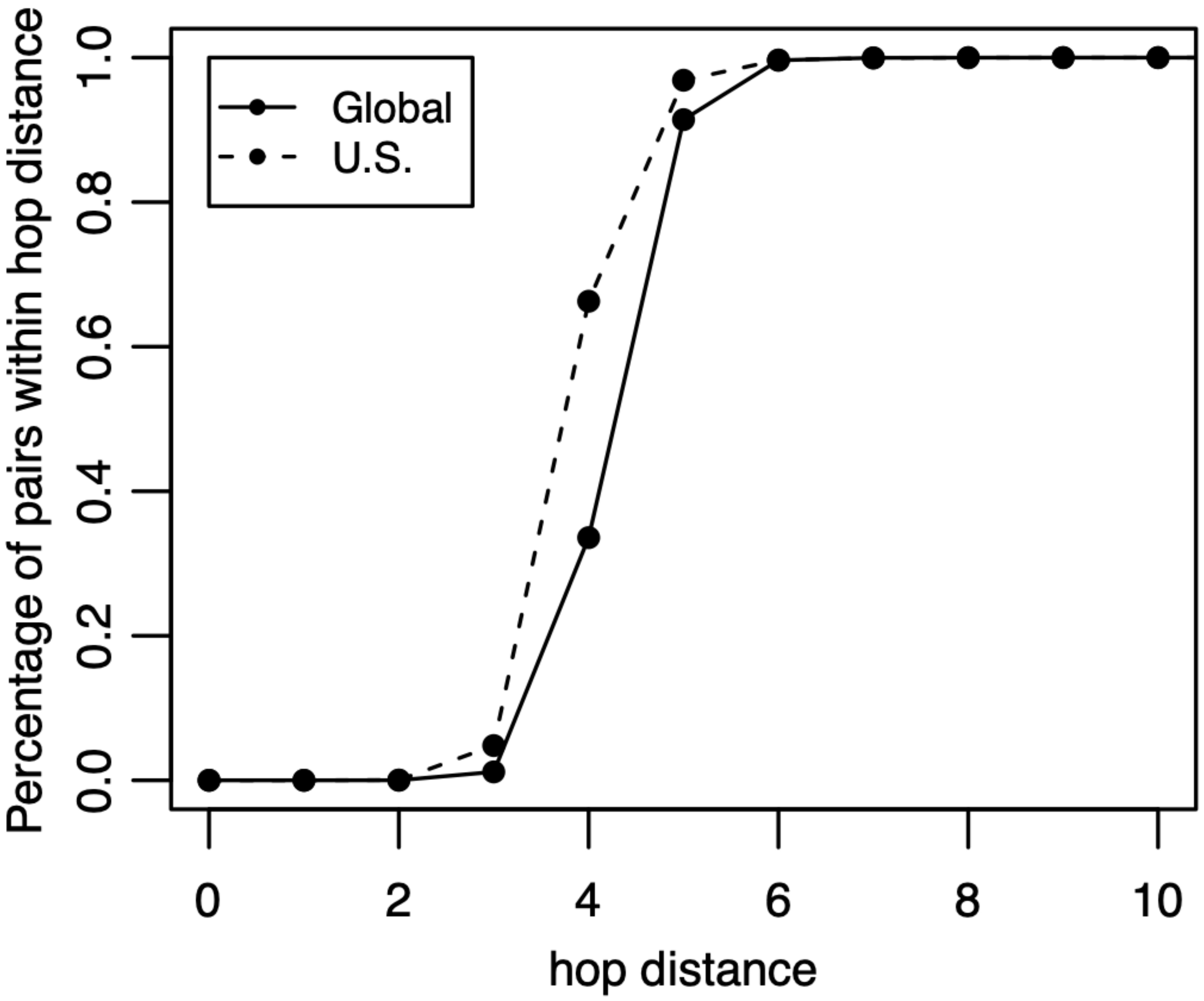
$$c_i = \frac{6}{21} \approx 0.29$$


What can we say about
individual nodes at the
context of the whole network?


What can we say about individual nodes at the context of the whole network?
Path length = minimum number of links one needs to follow to go from one node to another

Paths & small-worlds

- Path length: minimum number of links one needs to follow to go from one node to another
- Social networks are small worlds = low path lengths
- E.g., avg path length <4 in Facebook network with 721 million users

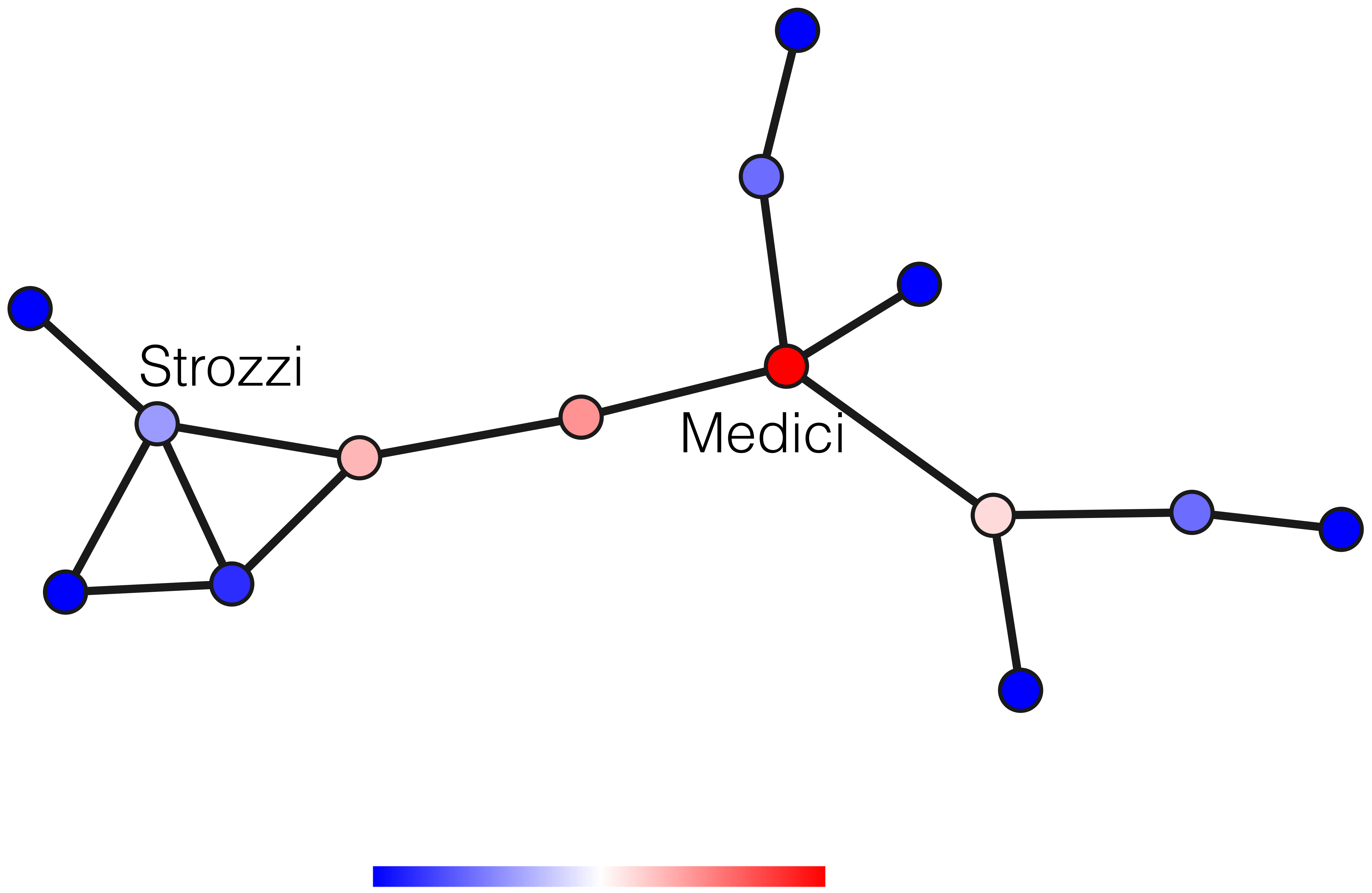



What can we say about individual nodes at the context of the whole network?
Can we quantify how “central” a node is?

Centrality

- Central nodes in social networks:
  - Control flow of information
  - Keep the network connected
  - Important for spreading phenomena

Betweenness centrality
=
How many shortest paths pass through a node?



Centrality measures

What makes a node central and important?

1. Degree centrality

Important nodes have many connections

2. Betweenness centrality

Important nodes work as bridges

3. Closeness centralities

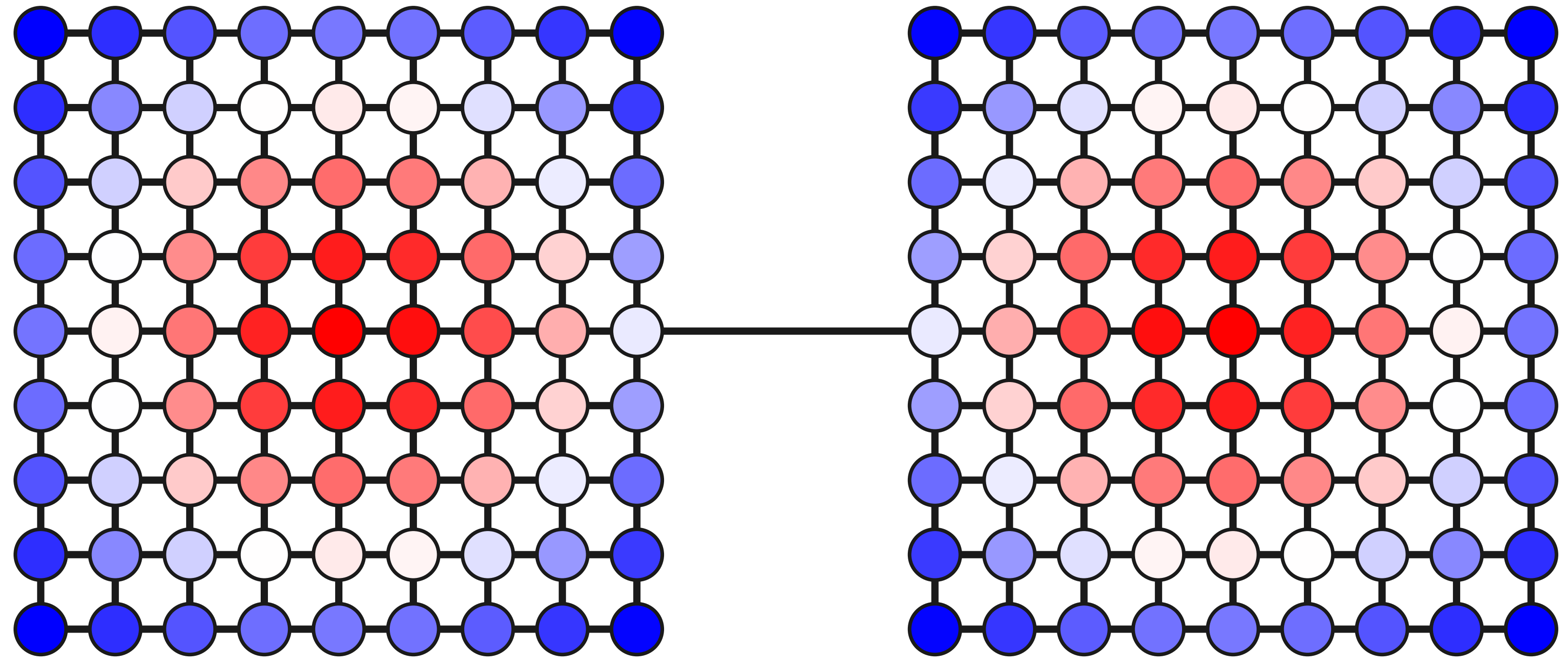
Important nodes are close to other nodes

4. Eigenvector centralities

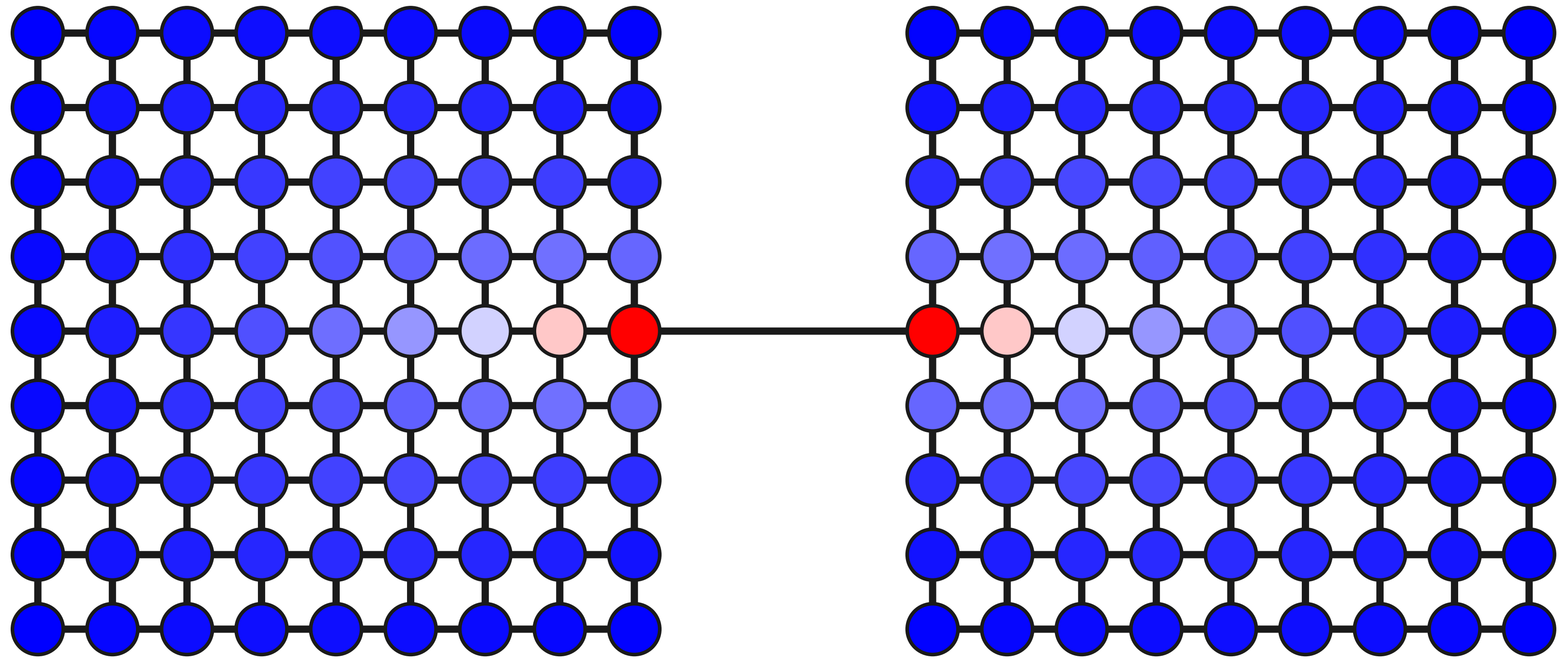
Important nodes are connected to other important nodes

Centralities in an example network

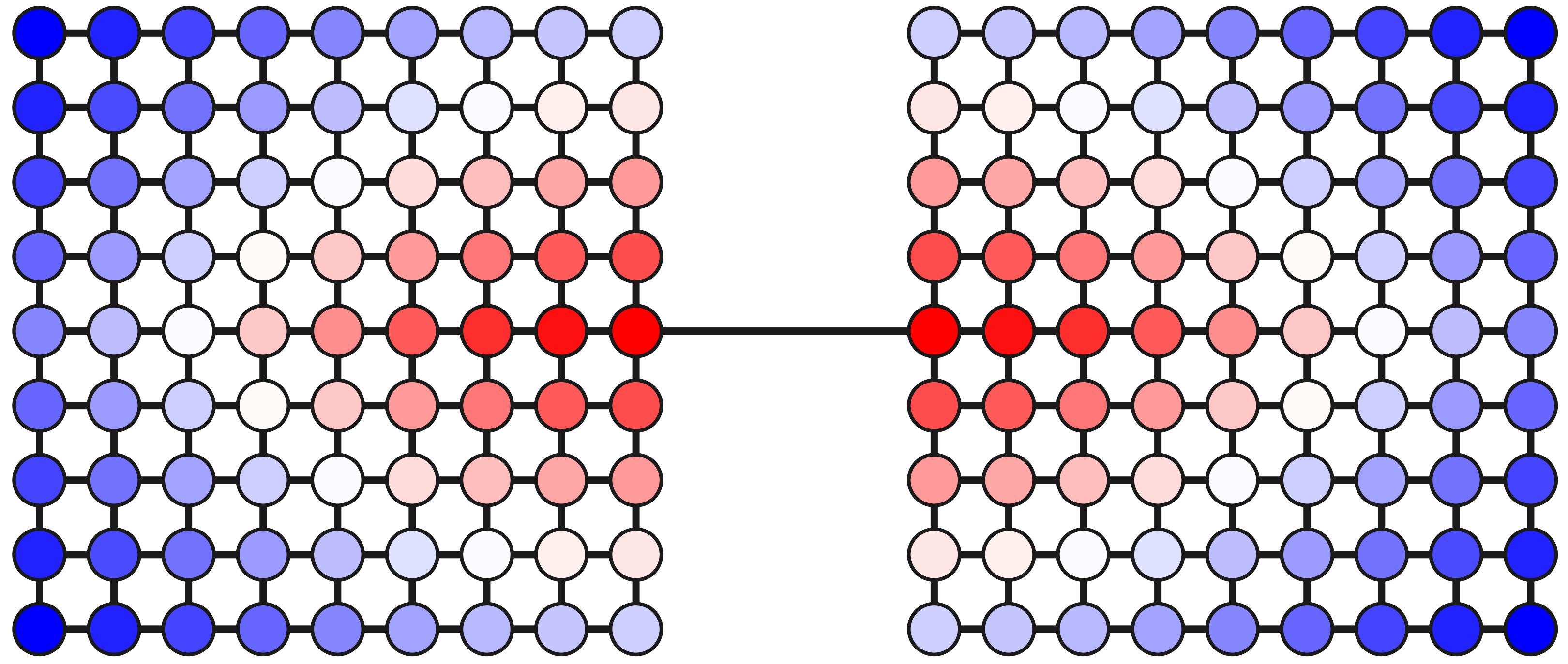
Eigenvector



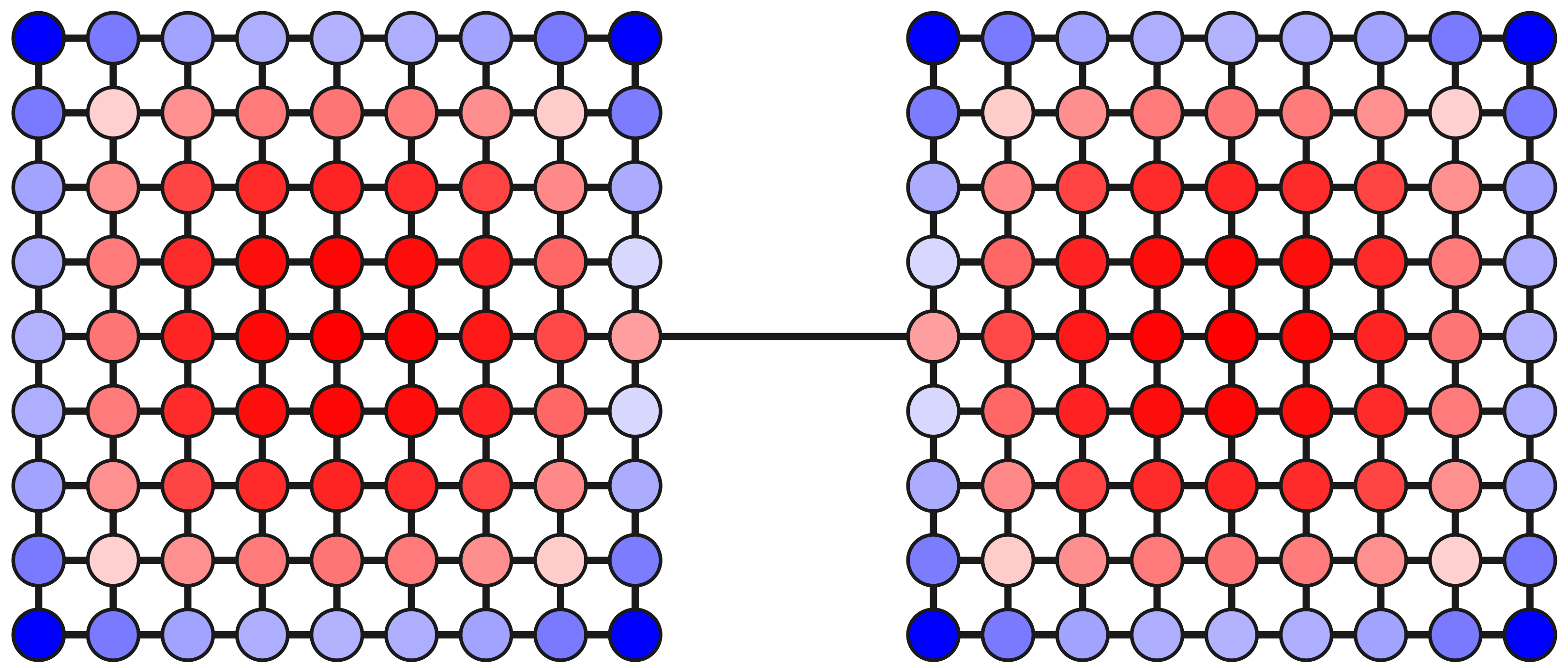
Betweenness



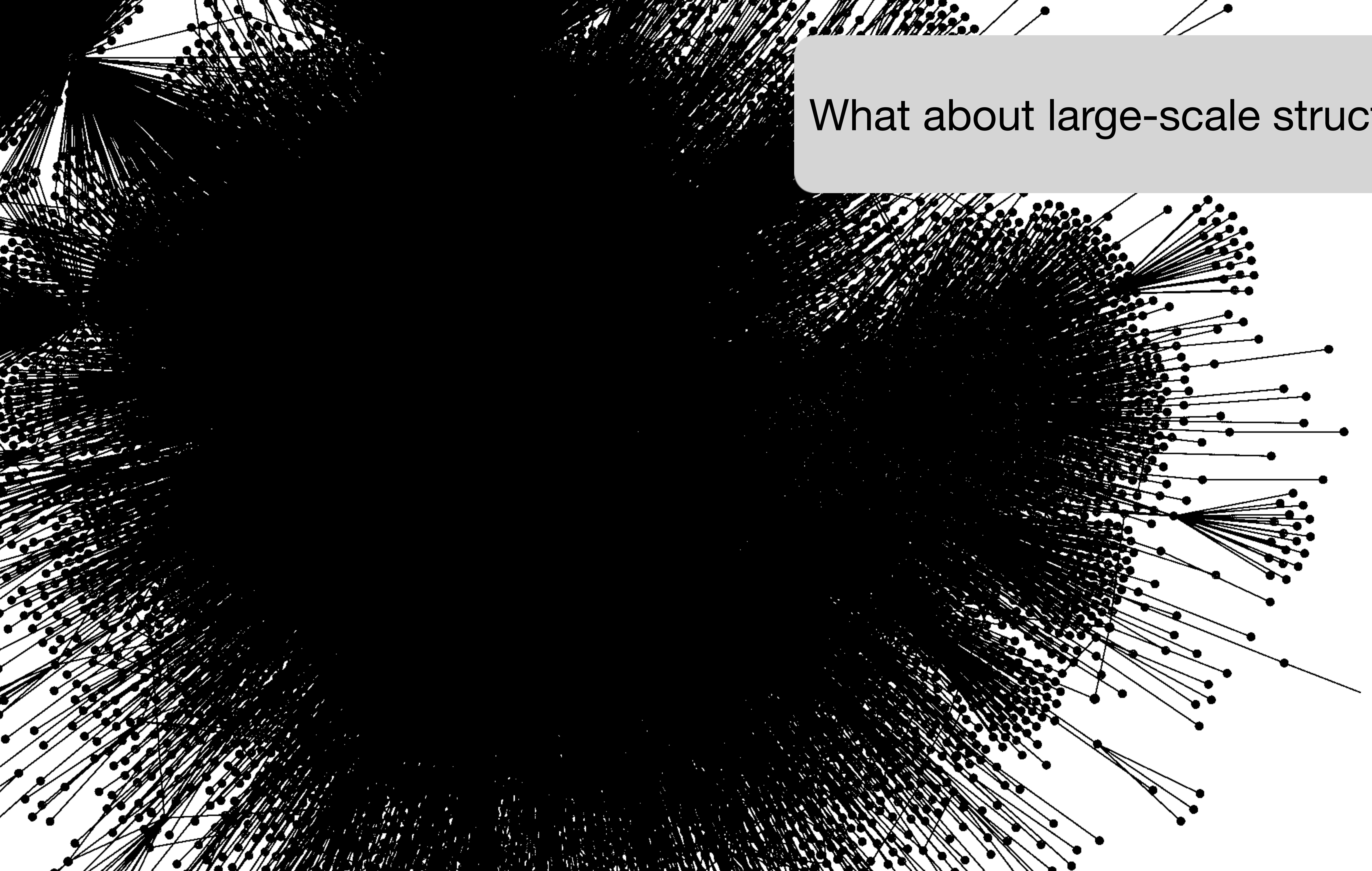
Closeness



Katz, a=0.2



Low value  High value

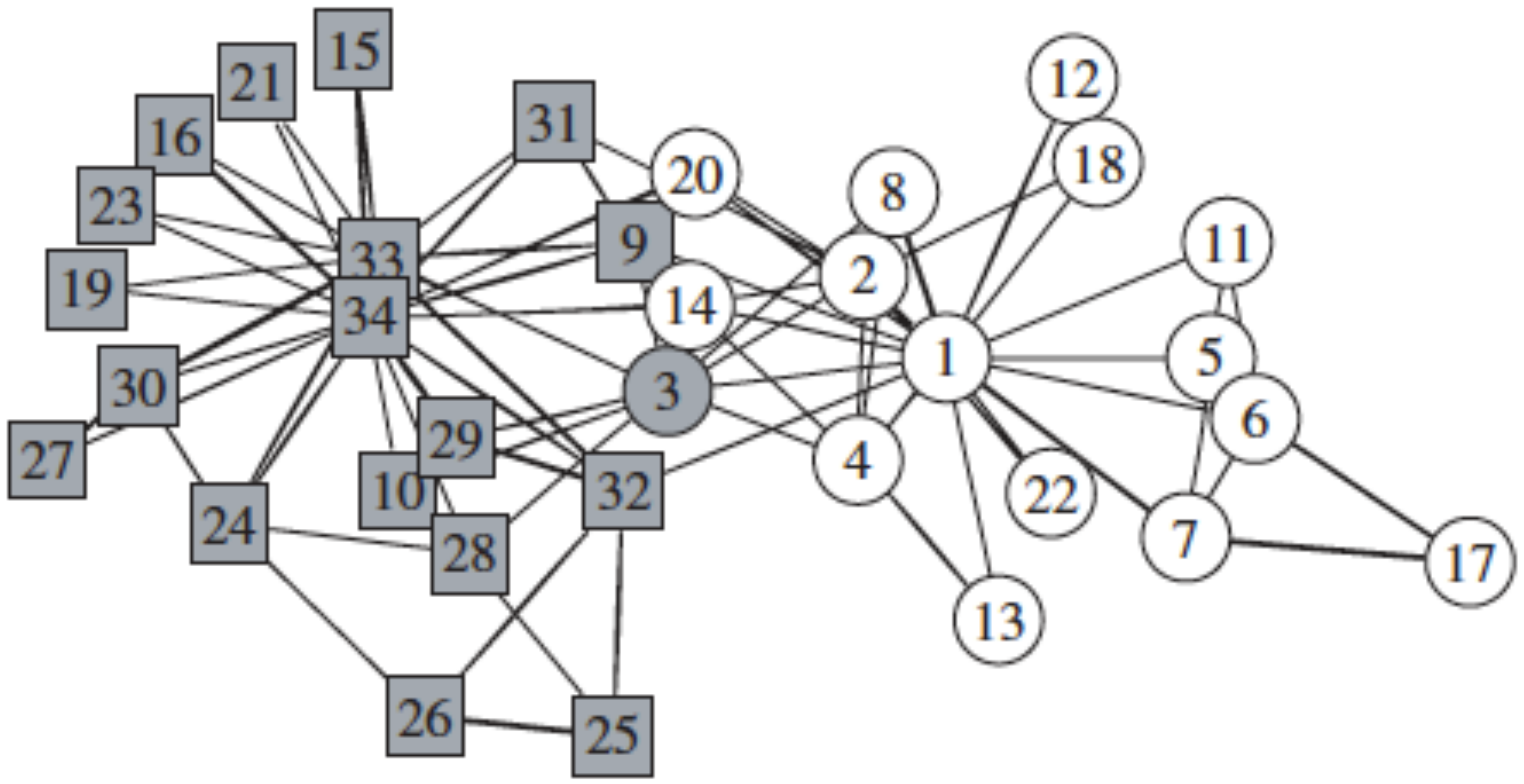


What about large-scale structures?

What about large-scale structures?

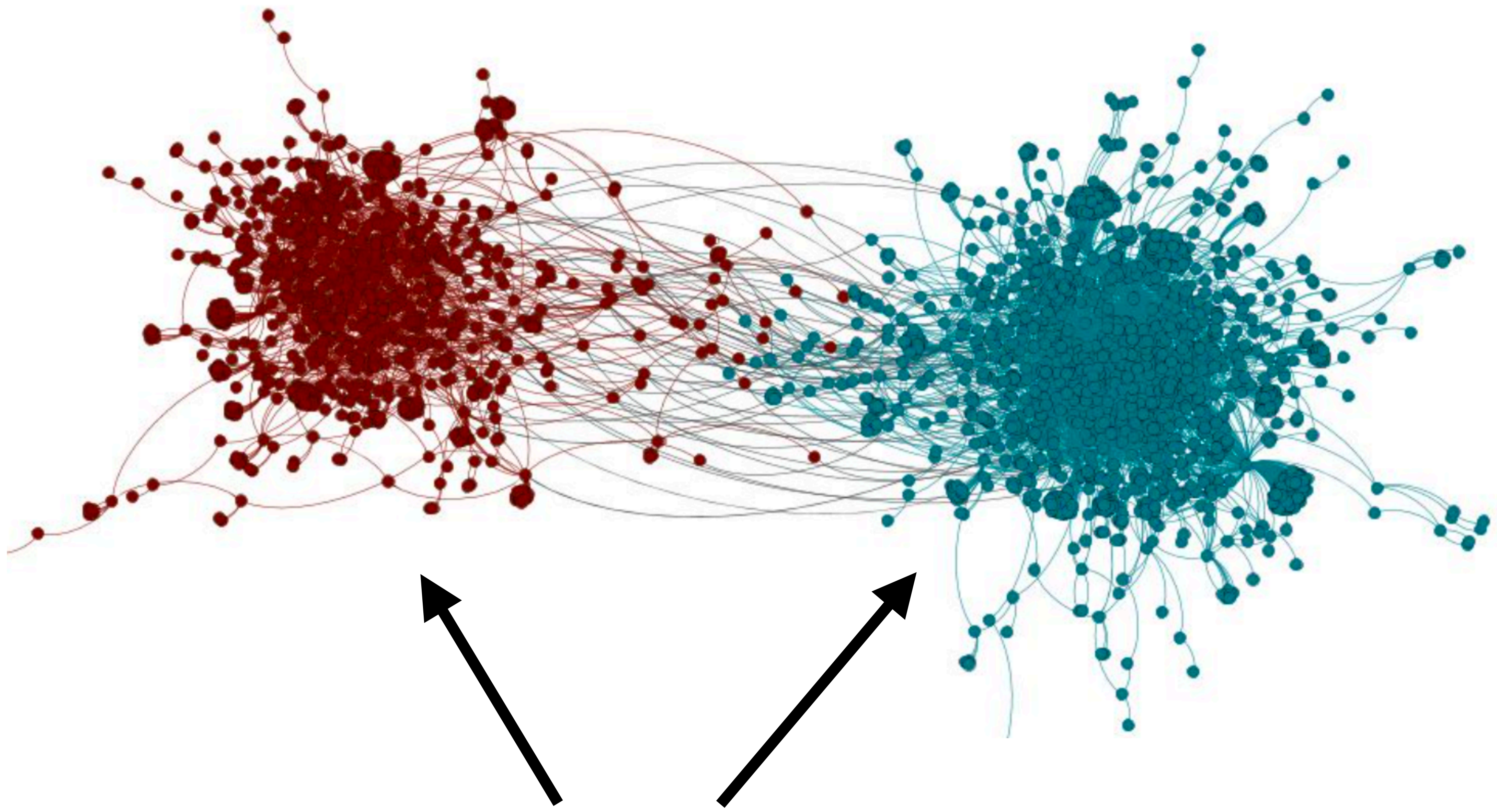
Communities and clusters

- Set of densely interconnected nodes
- Social groups
- Small (few nodes) - large (millions of nodes)
- Terminology: Community = cluster = module



Community detection methods

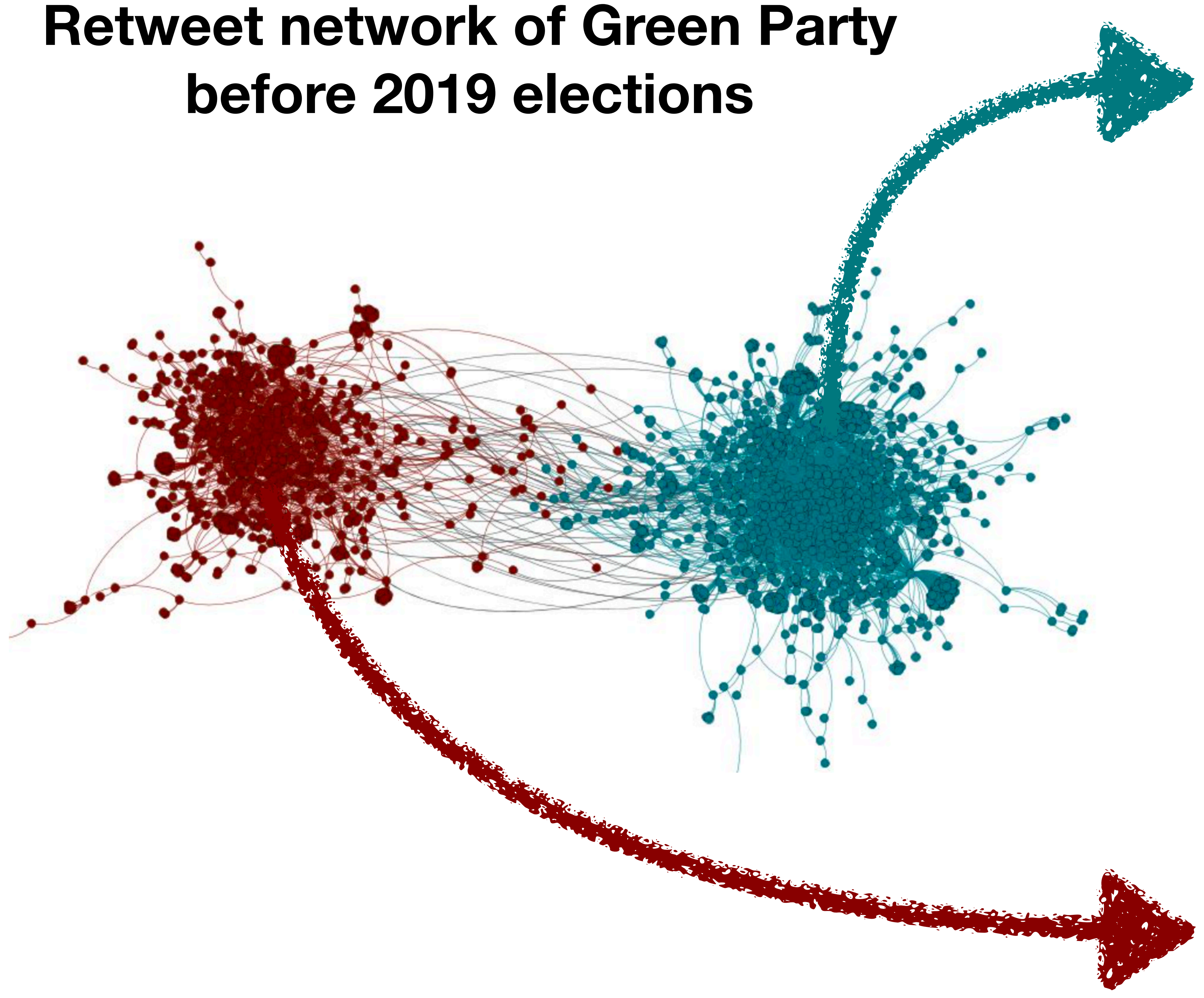
- Large number of formal definitions
  - Find communities = ill defined problem
  - Different methods can find different structures
- Difficult algorithmic problems
- Most methods subject to overfitting
- Some common methods: *modularity, informal, stochastic block models*



Colors: groups found with a graph clustering method

Example: clusters in social media

Retweet network of Green Party before 2019 elections



1. Me and many others will remember Sipilä's government from how they were silent in front of the far right, and how they agreed to inhumane refugee politics. That is why I'm #green, we do not remain silent and don't accept inhumane politics. #antiracism #greens
2. Our generation can still control the climate change. Our children will not have that possibility. Voting is now more important than ever. In my video I will tell what kind of Finland I want to leave for the next generations. What about you? #elections2019 #greens [url]
3. Greens are becoming the largest party in the capital, think about that! Amazing! #greens #elections2019

1. Parliament members drove around with taxis during the last term with 1.8 million EUR. The Green Party is leading the statistics. Ville Niinistö doesn't feel like walking even 700 meter distances! #politics #greens [url]

2. Children are forced vegan, cars are taken away from people, and energy bills keep rising. #greens Jobs for foreigners and everything is privatised. #nationalcoalitionparty Finland should think of Finns. #truefinns Which one do you choose?

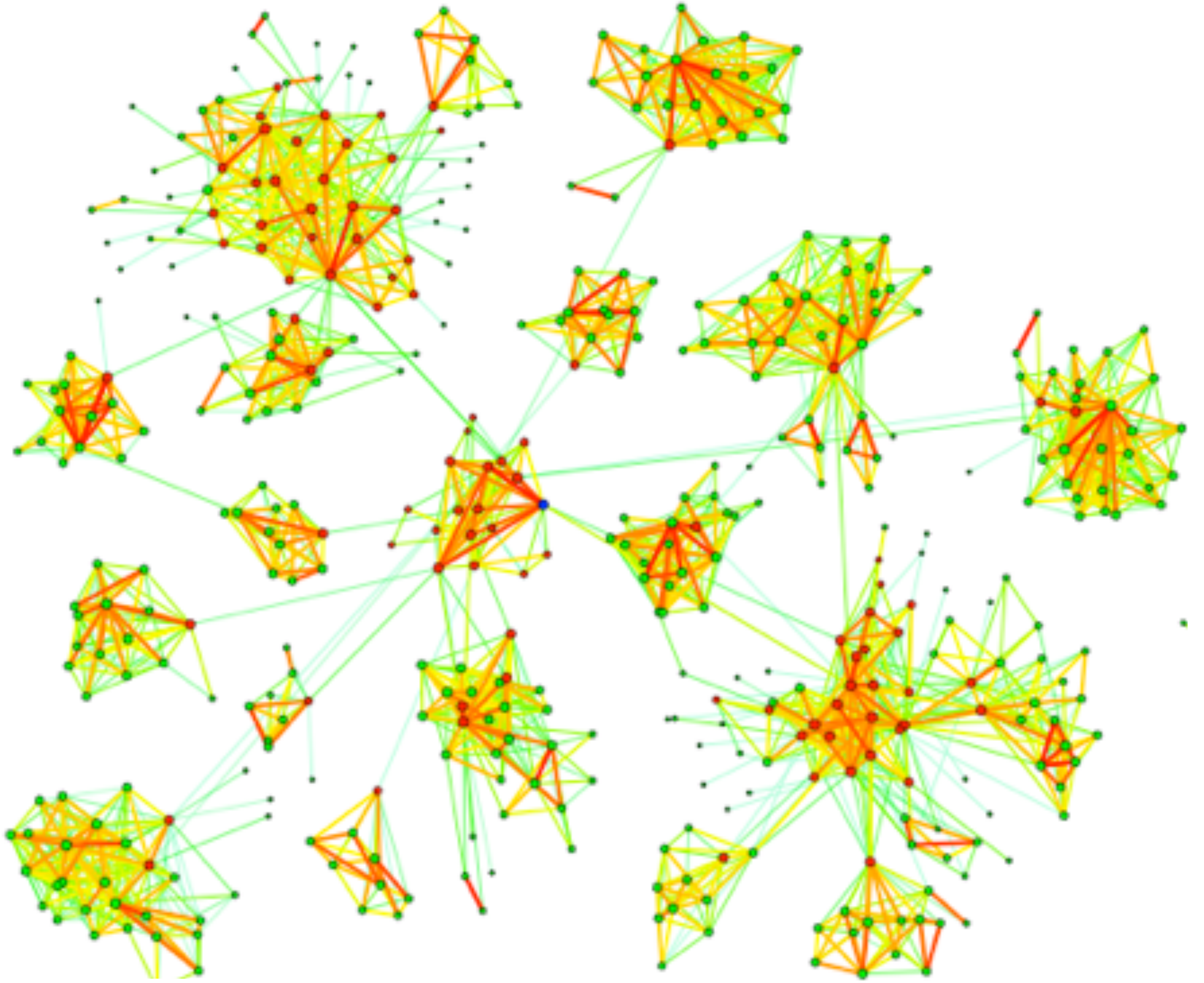
3. Hypocritical #greens don't care about electrical cars or biodiesel as fuel, but instead they just drive around with the same petrol and diesel cars as normal people. Don't do as we do, but do as we preach. #elections2019 #climate [url]

Groups and tie strengths

- “Strength of weak ties”: weak links connect clusters - strong links are inside clusters

→ Networks held together by weak links; weak links bottlenecks for information spreading

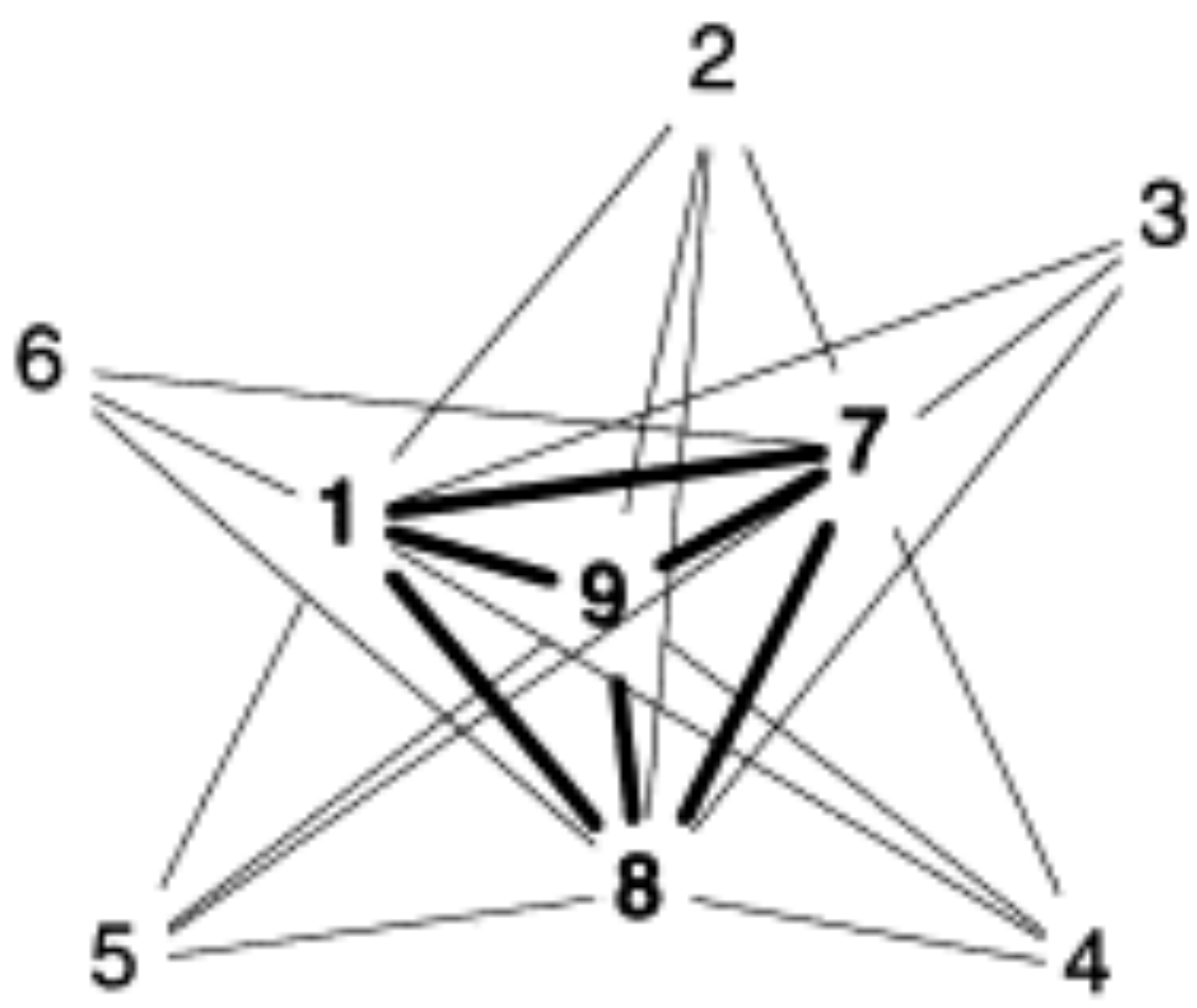
- Original theory in 1973: later proved with country-scale data



Granovetter, M. S. (1973). The strength of weak ties. *American journal of sociology*, *78*(6), 1360-1380.

Cores and peripheries

- Core: highly internally connected
- Periphery: sparsely internally connected
- People in the core in a more powerful position



Borgatti, S. P., & Everett, M. G. (2000). Models of core/periphery structures. *Social networks*, *21*(4), 375-395.

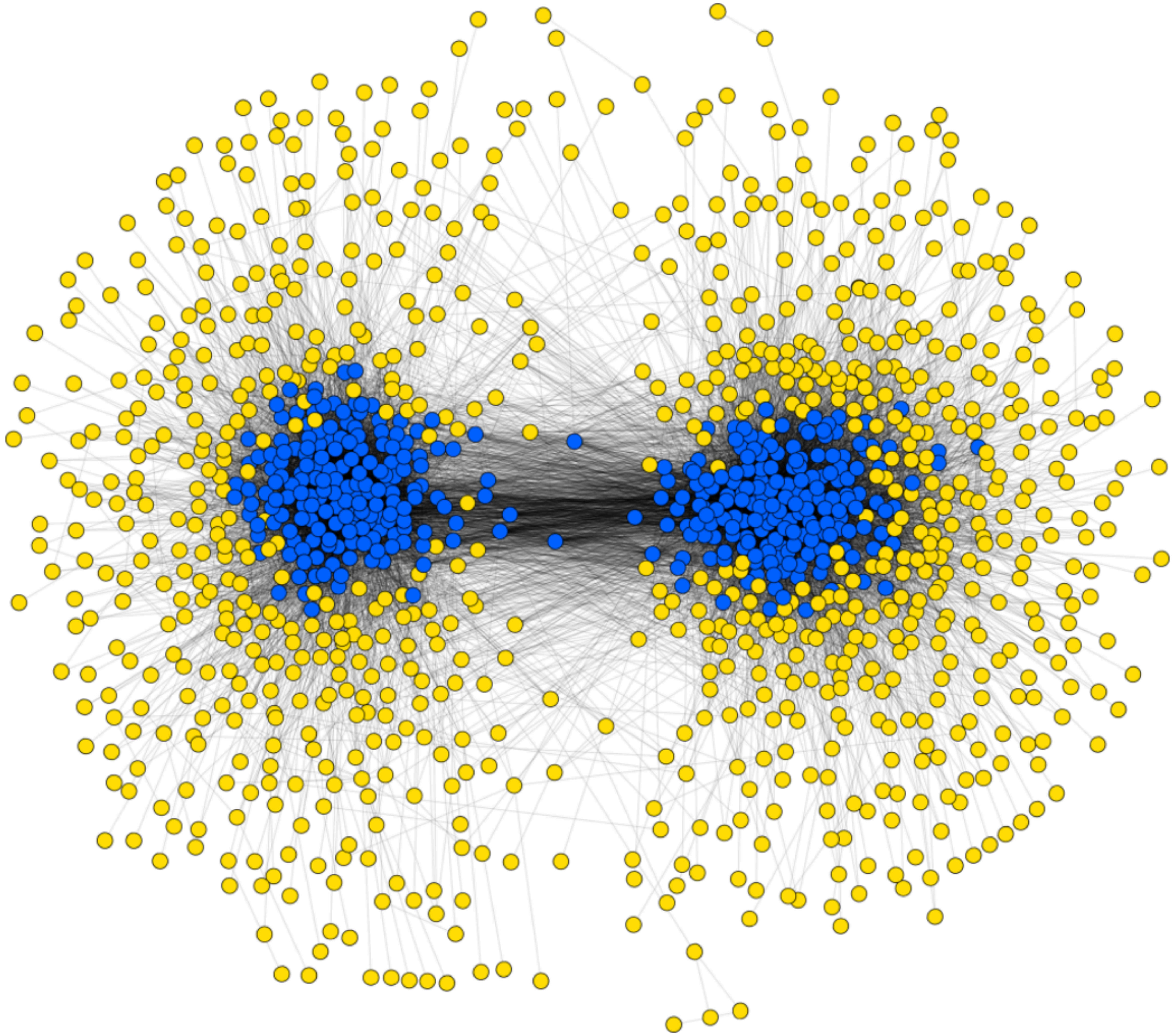
Other types of coarse grained structures

- More complicated "mesoscale" structures?

- Stochastic block model:

    - Divide the network into groups of nodes such that the nodes in each group are connected in similar ways to the groups

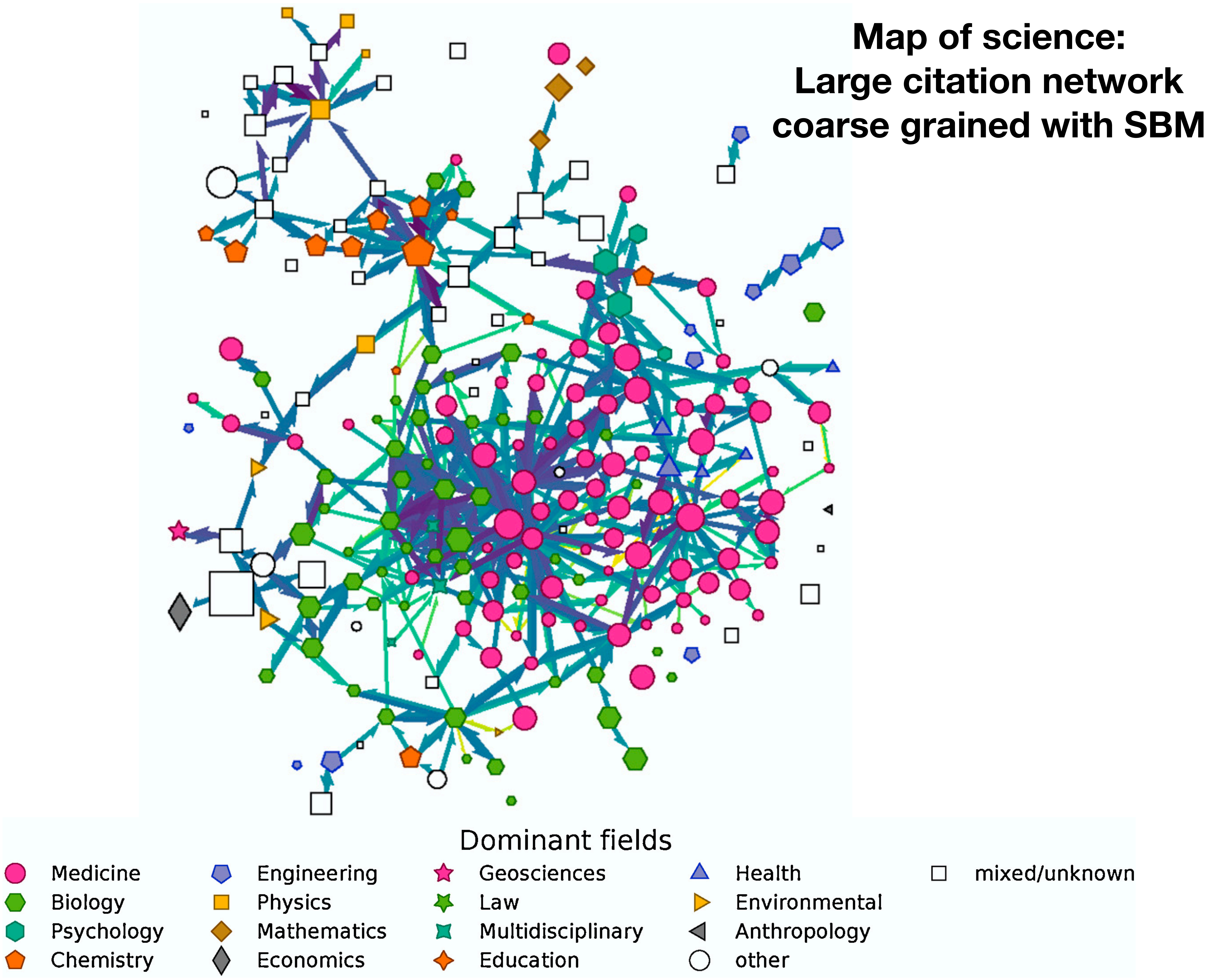
Political blogs:
2 core-periphery structures or 2 communities?



Zhang, X., Martin, T., & Newman, M. E. (2015). Identification of core-periphery structure in networks. *Physical Review E*, *91*(3), 032803.

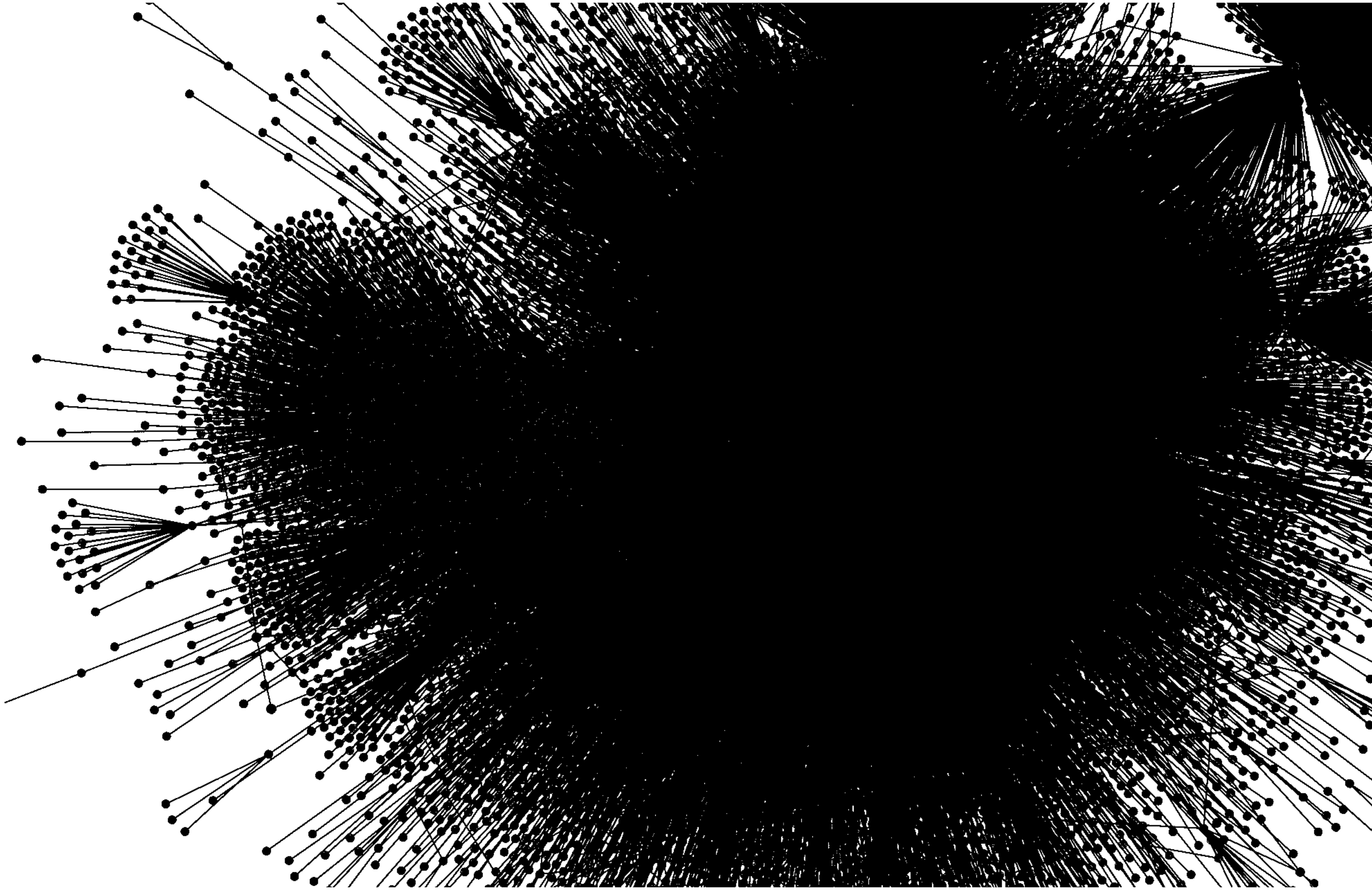
Other types of coarse grained structures

- More complicated "mesoscale" structures?
- Stochastic block model:
    - Divide the network into groups of nodes such that the nodes in each group are connected in similar ways to the groups



Network visualization

- Visualization is a good starting point
- Too dense network → hairball
    - Network can still have interesting structure
- Visualization algorithms: keep linked nodes close to each other (but not overlapping too much)



What explains the emergence of the structure?

How does the structure affect dynamics?

What explains the emergence of the structure?

How does the structure affect dynamics?

Answers in lecture 9

Questions?

Midjourney:
https://creativecommons.org/licenses/by-nc/4.0/legalcode